A la recherche des désintégrations des mésons beaux vers des états finaux purement leptoniques

M. Perrin-Terrin

CPPM-T-2013-03

Aix-Marseille Université
Faculté des Sciences de Luminy
163 avenue de Luminy
13288 MARSEILLE Cedex 09

Thèse de Doctorat

Spécialité : *Physique et Sciences de la Matière*

Mention : *Physique des Particules et Astroparticules*

présentée par

Mathieu Perrin-Terrin

en vue d'obtenir le grade de docteur de l'Université d'Aix-Marseille

À la recherche des désintégrations des mésons beaux vers des états finaux purement leptoniques

Searches for B meson decays to purely leptonic final states

soutenue le 23 septembre 2013 devant le jury composé de

Dr.	S. Descotes-Genon	Examinateur
Prof.	T. Gershon	Rapporteur
Dr.	G. Hamel de Monchenault	Rapporteur
Dr.	E. Kajfasz	Examinateur
Dr.	R. Le Gac	Co-directeur de thèse
Dr.	G. Mancinelli	Directeur de thèse

Remerciements

Au moment d'écrire ces remerciements, j'ai soutenu ma thèse et quitté la collaboration LHCb pour rejoindre NA62 au CERN. Ces deux événements m'ont fait comprendre la phrase naïve que j'entendais souvent :

la physique des particules est une grande famille.

Lors de ma soutenance de thèse, j'ai eu la chance non seulement d'avoir un jury qui a su guider la séance de questions vers une vraie discussion scientifique mais qui a aussi voulu laisser s'exprimer son empathie et son implication dans ce grand moment de ma vie. Je tiens à leur exprimer ma plus grande gratitude pour avoir permis à ce moment scientifique d'être aussi un moment humain, et, en particulier, d'avoir été témoins, avec toute l'audience, de l'expression des liens quasi filiaux entre un directeur de thèse et son étudiant.

Arrivé au CERN, j'ai eu la chance de pouvoir choisir dans quelle expérience travailler, et guidé par la curiosité scientifique j'ai décidé de m'engager dans NA62. Mais au moment d'officialiser ce choix, qui signifiait aussi quitter LHCb, j'ai pu mesurer, à mes dépends, l'attachement que j'avais pour tous ces collaborateurs de LHCb. Je me suis senti appartenir à une famille.

Les phrases précédentes laissent déjà transparaître la relation très forte que nous avons construit Giampiero et moi durant cette thèse. Je ne pourrais jamais le remercier pour tout ce qu'il m'a donné. Il a fait de moi un apprenti physicien mais j'ai aussi beaucoup appris de ses qualités humaines hors du commun.

Il y a cinq ans, engagé dans des études d'ingénieur, je cherchais à me réorienter vers de la physique plus fondamentale. Giampiero me donna alors une chance et je passais trois mois en stage avec lui. Un an après, je voulais revenir faire ma thèse. Il a alors remué ciel et terre (Justine peut temoigner) pour que j'obtienne une bourse. Trois ans après, c'était encore lui qui me poussait à postuler pour ce fellowship au CERN.

Tout au long de ma thèse il a su me faire travailler en comprenant la personne que je suis. Un tel encadrement m'a permis de m'épanouir dans

mon travail de thèse et de mieux connaître mes forces et mes faiblesses, mes besoins et mes limites. Peut être la meilleur façon de le remercier est de lui dire que j'espère un jour pouvoir transmettre autant à quelqu'un.

Je tiens également à remercier tous les membres du groupe LHCb de Marseille. Justine qui m'a suivi dès mes débuts à Marseille. Ses bons conseils ainsi que son aide technique avec le logiciel m'ont été précieux. Je lui suis aussi reconnaissant de n'avoir jamais hésité à se mettre en arrière plan pour contribuer à rendre mon travail – ou celui de n'importe quel etudiant – visible dans la collaboration. Je remercie aussi ceux avec qui j'ai partage le bureau 306, Cosme et Émilie d'abord puis Alessandro et Walaa ensuite. Bon courage les petits pour finir la thèse ! J'ai beaucoup aimé discuter avec Olivier et apprendre de sa grande expertise de la violation de CP. J'ai eu la chance de travailler avec Julien et de bénéficier de son discernement et de sa méthodologie tous les deux exceptionnels. Je remercie Renaud de m'avoir accueilli dans le groupe LHCb qu'il dirigeait. Je suis aussi très reconnaissant envers Elie pour ses fins conseils et sa confiance. Je n'oublie pas Andrei pour son aide précieuse au moment lancer sur la grille les calculs de limites - évidemment toujours à la dernière minute !

Je pense aussi à beaucoup de membres de la collaboration LHCb avec qui j'ai interagi presque quotidiennement notamment Marc-Olivier, Diego, Francesco, Gaia et Matteo. Je suis sûr que j'en oublie.

J'ai beaucoup apprécié l'environnement du laboratoire. Je remercie son directeur, Éric, notamment pour son écoute envers les étudiants. Je remercie aussi les services généraux ainsi que le service des missions que j'ai souvent sollicité. Et bien sûr, sans Thierry, et le support informatique impeccable qu'il assure, cette thèse aurait été plus compliquée.

Beaucoup de mes pensées vont vers mes amis. Ceux avec qui j'ai partagé mon quotidien Margaux, Nicolas, Nancy, Sophia. Mon colocataire et ami, Maxime, qui a été là tout le temps pour moi – dans les moments de déconne comme dans les moments plus sombres. Mes amis de longue date qui malgré l'éloignement m'ont beaucoup soutenu Matthias, Audrey, Yolaine, Clément, Julie. Et bien sûr Marion sans qui cette dernière année de thèse aurait été plus dure. Marion, je te remercie du fond du coeur pour ton écoute, ton immense compréhension et tes conseils toujours justes. Enfin, j'ai un tendre souvenir pour Sarah.

Des livres m'ont aussi accompagné durant ces trois années. J'ai en particulier trouvé beaucoup d'inspiration, et de secours parfois, dans les Mémoires d'Hadrien de Marguerite Yourcenar.

Je remercie mes frères, Pierre, Renaud et Jean-Baptiste qui m'ont beaucoup soutenu dans l'écriture de ce manuscrit. Pour finir, je tiens à remercier infiniment mes parents. Leur soutien a toujours été inconditionnel et ils ont

su m'aiguiller vers les choix les meilleurs pour moi. C'est d'ailleurs en discutant avec eux dans un petit bistrot à Londres que j'ai pris la décision de m'engager dans cette thèse.

Genève, 4 Décembre 2013

Contexte Théorique

Le Model Standard de la physique des particules (MS) est la théorie utilisée pour décrire la matière à l'échelle microscopique. Les prédictions de cette théorie ont été vérifiées expérimentalement avec une grande précision. Néanmoins, plusieurs observations – principalement d'origine astronomique – ne sont pas décrites par le MS et des problèmes internes mathématiques entachent la théorie. Ainsi, malgré son succès, le MS n'est probablement qu'une version effective à basse énergie d'une plus vaste théorie. La recherche de cette physique au delà du MS constitue le cœur du programme scientifique international de la physique des particules décrit récemment en details à la fois par le Colloque SnowMass aux Etats-Unis et par le Groupe Européen de Stratégie pour la Physique des Particules [1, 2]. Plusieurs voies d'entrée vers cette nouvelle physique ont été identifiées dont les désintégrations très rares des mésons B vers les états finaux purement leptoniques.

Ces désintégrations se produisent par l'intermédiaire de courants neutres changeant la saveur (CNCS), $\mathsf{b} \to \mathsf{s}\ell^+\ell^-$ dans le cas présent. Dans le MS, la saveur est changée au premier ordre uniquement par des courants charges (CCCS) et les CNCS ne se produisent qu'aux ordres supérieurs. Ainsi le MS prédit une très faible probabilité pour ces désintégrations. En outre, ces CNCS se font au travers de particules hors de leur couche de masse. Ainsi de nouvelles particules – avec des masses très élevées– peuvent intervenir dans ces courants et potentiellement modifier sensiblement l'amplitude de la désintégration par rapport à la faible valeur prédite par le MS. Les désintégrations $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ et $\mathsf{B}^0 \to \mu^+\mu^-$ sont d'autant plus sensibles à ces hypothétiques nouvelles particules que les contributions du MS sont supprimées d'hélicidé. Enfin les erreurs théoriques sur la prédiction des rapports d'embranchement de ces modes sont faibles et bien contrôlees, ainsi toute déviation de plus de 10% entre les rapports d'embranchement mesurés et ceux prédits par le MS,

$$\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-) = (3.57 \pm 0.30) \times 10^{-9}, \quad (1)$$
$$\mathcal{B}(\mathsf{B}^0 \to \mu^+\mu^-) = (1.07 \pm 0.10) \times 10^{-10}, \quad (2)$$
$$\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \tau^+\tau^-) = (7.57 \pm 0.63) \times 10^{-7}, \quad (3)$$
$$\mathcal{B}(\mathsf{B}^0 \to \tau^+\tau^-) = (2.24 \pm 0.10) \times 10^{-8}, \quad (4)$$

serait un signe indiscutable de nouvelle physique. La sensibilité de ces modes à la nouvelle physique – illustrée par la Figure 1 – est donc unique et a engendré des efforts expérimentaux considérables depuis 1984 [3] pour mesurer leur rapport d'embranchement. Les travaux développés dans cette thèse ont contribué à l'établissement de la première évidence de la désintégration

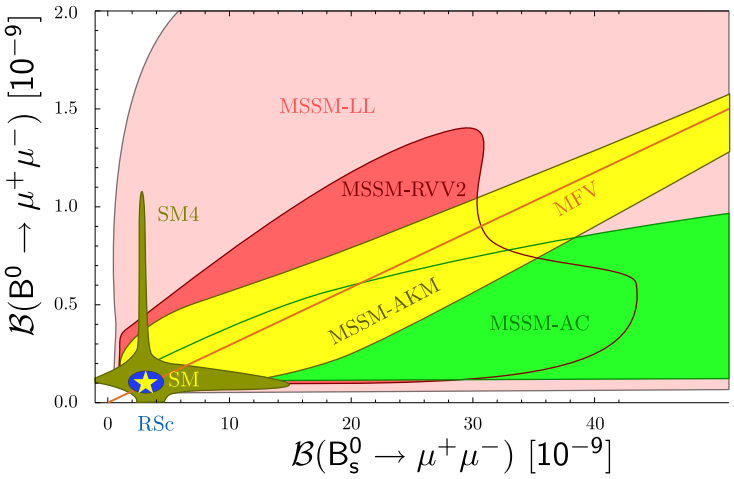


Figure 1: Prédictions pour les rapports d'embranchement de $B_s^0 \to \mu^+\mu^-$ et $B^0 \to \mu^+\mu^-$ dans les modèles à violation minimale de la saveur (MFV), le MS avec une quatrième génération (SM4), le modèle Randall-Sundrum (RSc) avec protection custodiale [4], quatre modèles minimaux de supersymétrie (MSSM), Agashe and Carone (AC) [5], Ross, Velasco-Sevilla and Vives (RVV2) [6], Antusch, King and Malinsky (AKM10) [7] et un modèle avec des courants gauches seulement (LL) [8]. La prédiction du MS est représentée par l'étoile (modifié de [9]).

$B_s^0 \to \mu^+\mu^-$ ce qui marque une étape majeure dans l'histoire de ces recherches. Ils ont aussi permis une interprétation phénoménologique plus fine de ces résultats.

Première Evidence de $B_s^0 \to \mu^+\mu^-$ à LHCb

La recherche des désintégrations $B_{(s)}^0 \to \mu^+\mu^-$ a été l'une des motivations à la construction d'un détecteur dédié à la physique des saveurs lourdes auprès du LHC. Une vue schématique de ce détecteur, LHCb, est présentée dans la Figure 2. La conception du détecteur vise à exploiter au maximum l'environnement du LHC pour la physique du B. Les collisions proton-proton (pp) avec une énergie au centre de masse de plusieurs TeV produisent, comme le montre la Figure 3, des mésons B faisant des petit angles avec le tube de faisceau. Ainsi le détecteur doit d'instrumenter ces régions, et plutôt que diviser les ressources financières pour instrumenter les deux régions (vers l'avant et vers l'arrière), il a été choisi d'équiper seulement une région avec des dispositifs plus performants. Le detecteur LHCb est donc est spectromètre à un seul

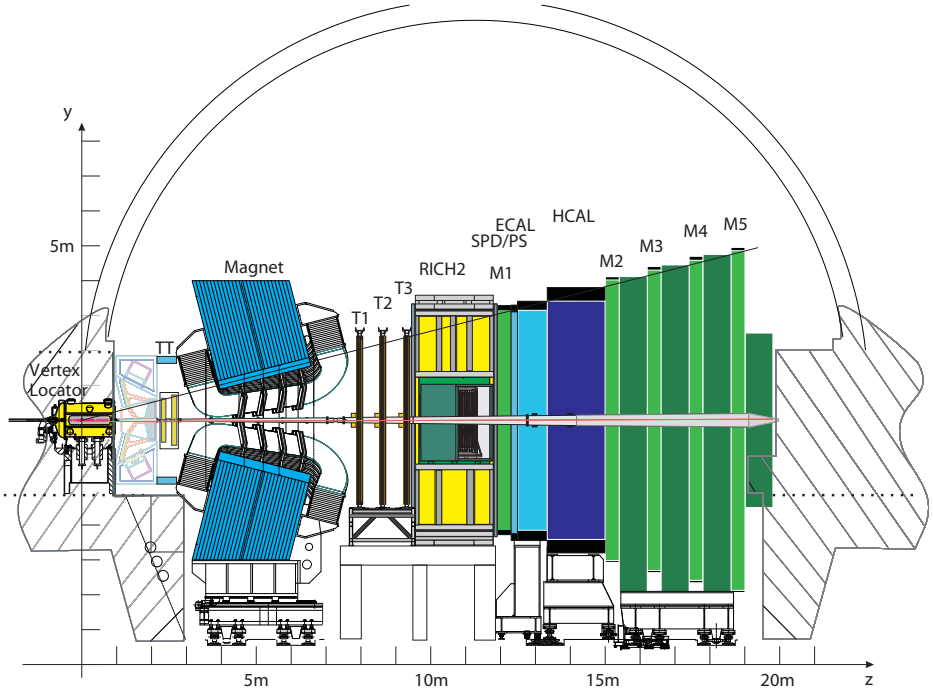


Figure 2: Coupe du détecteur LHCb.

bras, dirigé vers l'avant. Le détecteur se compose de sous-systèmes visant à remplir trois fonctions principales : la reconstruction des traces et des vertex, l'identification des particules et enfin la sélection ultra rapide des évènements intéressants. Cette dernière fonction permet de réduire le flot d'évènements produit par les collisions pp du LHC à 40 MHz à un niveau gérable de quelques dizaines de kHz. Dans le cas de la désintégration $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$, ces sous-système atteignent des performances excellentes. Par exemple la résolution sur le moment des B est meilleure que 0.6% à 500 GeV/c, celle sur le paramètre d'impact des muons avec un moment transverse de 2 GeV/c est meilleure que 25 μm, et le taux d'hadrons (K^+, π^+) identifies comme des muons est inferieur à 0.6% pour un taux d'identification correcte des muons de 98%. Avec de telles performances, LHCb fournit un environnement idéal pour la recherche des désintégrations $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ et $\mathsf{B}^0 \to \mu^+\mu^-$.

La stratégie pour analyser les données, enregistrées avec le détecteur LHCb, se décompose en quatre étapes principales (sélection, classification, normalisation et calcul des limites sur le rapport d'embranchement ou leur mesure) et s'appuie autant que possible sur des canaux de control plutôt

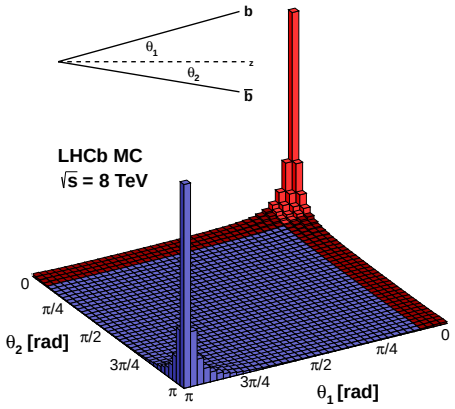


Figure 3: Corrélation angulaire entre les b et $\overline{\mathsf{b}}$ produits au LHC selon les données simulées [10, 11].

que sur des simulations. Pour éviter les biais inconscients, tous les choix de l'analyse sont faits sans regarder la région de masse invariante où se trouve le signal. Tout d'abord, une sélection multivariée lâche est appliquée pour réduire les différentes contributions de bruits de fonds suivants :

candidats combinatoires prompts, pour lesquels au moins un des deux muons formant le candidat provient du vertex primaire,

candidats combinatoires à grand temps de vie, pour lesquels le μ^- et le μ^+ proviennent des désintégrations semi-leptoniques du $\overline{\mathsf{b}}$ et du b

candidats B issus de cascades, $(\mathsf{B} \to \mathsf{D}(\to \mu\mathsf{X})\mu\mathsf{X})$, où l'un des muons provient de la désintégration du B en un méson charmé qui se désintègre à son tour et fournit le second muon pour former le candidat,

candidats $\mathsf{B}^0 \to \mathsf{h}^+\mathsf{h}'^-$ incorrectement identifiés, pour lesquels les deux hadrons se désintègrent en vol en un muon et un neutrino et sont alors identifiés comme des muons,

candidats provenant d'autres sources de désintégration exclusives, qui regroupent principalement $\mathsf{B}^{0,+} \to \pi^{0,+}\mu^+\mu^-$ et $\mathsf{B}^0 \to \pi^-\mu^+\bar{\nu}_\mu$ avec un pion identifié incorrectement ou les paires di-muons produites par le processus $\mathsf{p}^+\mathsf{p}^+ \to \mathsf{p}^+\mu^-\mu^-\mathsf{p}^+$.

Cette sélection est conçue pour être aussi applicable aux canaux de control. Après cette sélection, un discriminant multivarié (arbre de décision boosté) est entrainé à distinguer, grâce à des variables topologiques non corrélées avec la masse invariante, le signal du bruit de fond combinatoire, principale source de fond restante après la sélection. La sortie de ce discriminant est représentée en Figure 4 pour les données simulées du signal, du bruit de fond.

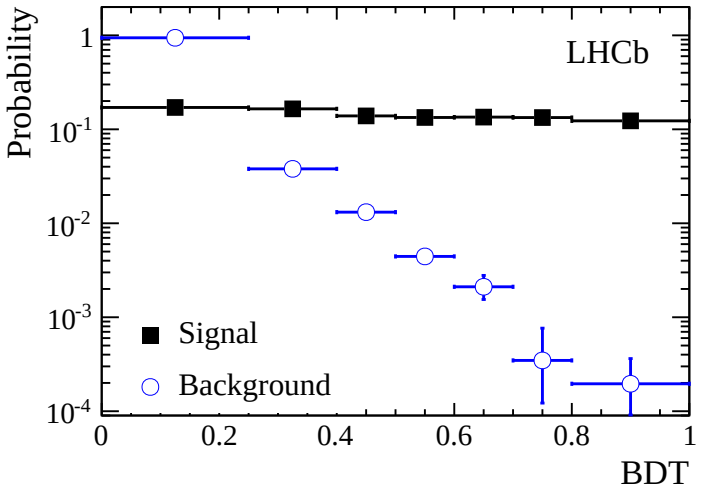


Figure 4: Fonctions de densité de probabilité pour le signal (carrés noirs) et bruit de fond (cercles bleus) obtenues à partir canaux de control.

Il est préférable, plutôt que d'utiliser ce discriminant multivarié pour faire une nouvelle sélection, de s'en servir avec la masse invariante pour construire des catégories et de garder ainsi tous les évènements de signal. Le choix de ces catégories est déterminant pour la sensibilité globale de l'analyse. Ce choix a été fait avec une méthode basée directement sur le test de statistique utilisé en aval pour calculer les limites sur les rapports d'embranchement. La méthode s'appuie sur un ajustement en deux dimensions des distributions de la masse invariante et du discriminant multivarié qui permet d'evaluer analytiquement le contenu en signal et en bruit de fond de n'importe quel ensemble de catégories. Ces estimations sont utilisées ensuite pour évaluer la signifiance avec laquelle un signal de type MS exclurait l'hypothèse d'un échantillon de bruit de fond pur. Plusieurs milliers d'ensemble de catégories ont été testées en imposant certaines contraintes nécessaires pour estimer le contenu de ces catégories par des méthodes basées sur des échantillons de control. L'ensemble de catégories optimal obtenu apporte une amélioration de la sensibilité de l'analyse de 14% par rapport à la catégorisation précédente

en 4×6 bins réguliers pour la masse invariante et discriminant multivarié.

Apres la sélection et la classification des événements, le contenu des catégories doit être estimé précisément. Le bruit de fond attendu est obtenu en ajustant hors de la région de masse invariante du signal, une fonction de densité de probabilité décrivant les différentes contributions de fond, et en extrapolant ensuite dans la région de masse invariant du signal. La fonction de densité de probabilité de la distribution de masse invariante et du discriminant multivarié du signal est obtenue à partir de canaux de control. Le nombre total d'événement de signal correspondant à une hypothèse de rapport d'embranchement est quant à lui obtenu se normalisant à deux canaux, $\mathsf{B}^+ \to \mathsf{J}/\psi\mathsf{K}^+$ et $\mathsf{B}^0 \to \mathsf{K}^+\pi^-$. Les rapports d'embranchement de ceux-ci étant connus, il est possible d'obtenir le nombre de B^+ et de B^0 produits en mesurant les nombres d'événements pour chaque canal ainsi que les efficacités pour les sélectionner. Les nombres de B^+ et de B^0 sont enfin convertis en nombres de B^0_s ou B^0 en utilisant le ratio des probabilités qu'un quark b s'hadronise avec un quark $\bar{\mathsf{s}}$ plutôt qu'un quark $\bar{\mathsf{u}}$ ou $\bar{\mathsf{d}}$. Ce ratio est la source principale d'incertitude systématique. Il est mesuré dans les données de LHCb [12, 13].

A ce stade de l'analyse les distributions attendues de signal et de bruit de fond sont connues. Il est alors possible d'estimer la valeur du rapport d'embranchement de $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ et de $\mathsf{B}^0 \to \mu^+\mu^-$ à partir de laquelle un échantillon contenant du signal et du bruit de fond serait statistiquement incompatible avec l'hypothèse du bruit de fond pur. Ces limites vers le haut attendues sur le rapport d'embranchement permettent de mesurer la sensibilité de l'analyse. La confrontation d'un échantillon de données aux deux hypothèses, bruit de fond pur ou signal et de bruit de fond, se fait avec un test de statistique Q défini par:

$$Q = \prod \mathcal{P}(d_i, s_i + b_i)/\mathcal{P}(d_i, b_i) \ , \tag{5}$$

où $\mathcal{P}(d_i, s_i + b_i)$ est la probabilité pour que le nombre d'évènements observé (d_i) corresponde à une fluctuation poissonienne du nombre d'évènements attendu (s_i+b_i). Cette grandeur permet de classer, pour chaque hypothèse de rapport d'embranchement du signal, un échantillon suivant sa comptabilité avec la présence ou non de ce signal. Pour calibrer cette échelle, des pseudo-expériences sont générées sous l'hypothèse du bruit de fond pur et du bruit de fond et signal. La séparation entre les distributions du $-2\ln Q$ de ces deux ensembles de pseudo-expériences donne une représentation graphique de la sensibilité de l'analyse comme le montre la Figure 5.

Plus les distributions sont séparées et plus la sensibilité de l'analyse est grande (en passant, le choix des catégories mentionné précédemment

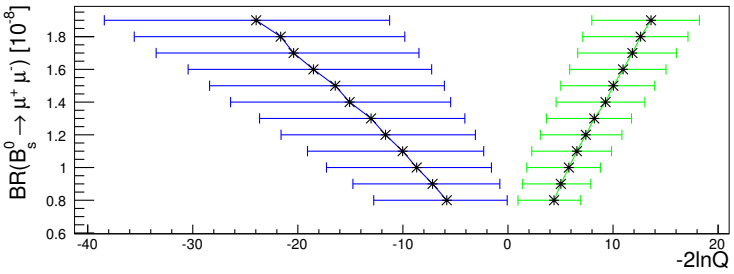


Figure 5: Valeur centrale du test de statistique des distribution generee pour differentes hypotheses du rapport d'embranchement du signal pour le bruit de fond pur (vert) et le bruit de fond avec le signal MS (bleu).

$$CLb = \int_{Q<Q_{data}} Q_b \mathrm{d}Q_b$$

$$CLsb = \int_{Q<Q_{data}} Q_{sb} \mathrm{d}Q_{sb}$$

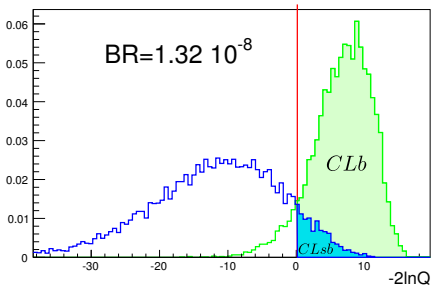


Figure 6: Définition du CL_{s+b} et du C.L.. Q_{data} est le test de statistique des données observées, Q_b et Q_{sb} les tests statistique sous les hypothèses signal plus bruit de fond et bruit de fond seulement.

a été optimisé sur la base de ce critère). Pour une expérience donnée, la compatibilité avec l'une ou l'autre des hypothèses est mesurée avec deux grandeurs, CL_{s+b} et CL_b définies sur la Figure 6 [14]. Une valeur de CL_b de 0.9973 ($1.0 - 5.7 \times 10^{-7}$) indique un signal avec une significance de 3 (5) σ. $1 - CL_{s+b}$ peut être utilisé comme niveau de confiance pour exclure des rapports d'embranchement. Néanmoins cette quantité peut amener à des limites abusives dans le cas, illustre sur la Figure 7, où l'analyse n'a aucune sensibilité au signal. Ainsi il est préférable d'utiliser $CL_s = CL_{s+b}/CL_b$ pour éviter ce cas de figure.

La pertinence statistique des limites et des signifiances obtenues avec cette méthode dépend grandement de la façon dont sont générées les pseudo-expériences. Ce processus de génération doit décrire correctement les fluctuations statistiques mais aussi les incertitudes systématiques sur les paramètres utilisés pour estimer les niveaux de signal et de bruit de fond. Un grand soin a été mis dans la description des densités de probabilités de ces paramètres

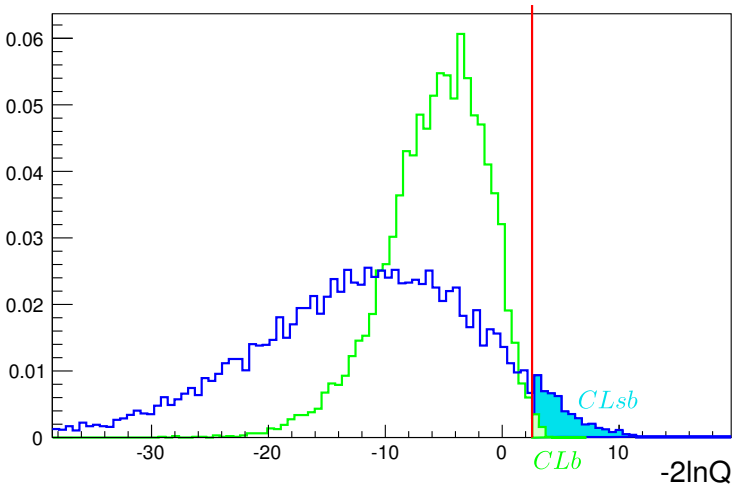


Figure 7: Cas de figure où, bien que l'analyse n'ait aucune sensibilité (large recouvrement des distributions du $-2\ln Q$ obtenues sous les hypothèses signal plus bruit de fond et bruit de fond pur), l'utilisation de $1-\mathrm{CL_{s+b}}$ comme indicateur de niveau de confiance amène à exclure abusivement des valeurs de rapports d'embranchement.

ainsi que leurs corrélations. Les quantités attendues de signal pour le MS et de bruit de fond dans les données obtenues avec les 1.1 et 1.0 fb^{-1} de collisions pp a 7 et 8 TeV sont reportées dans les trois premières lignes des Tables 1 et 2. En considérant les incertitudes statistiques et systématiques sur ces populations, un signal $\mathrm{B_s^0} \to \mu^+\mu^-$ MS est attendu avec une signifiance de 5.0σ et les rapports d'embranchement de $\mathrm{B^0} \to \mu^+\mu^-$ supérieurs à 7.1×10^{-10} pourraient être exclus. Les données observées sont reportées dans les dernières lignes des Tables 1 et 2. Une évidence de signal $\mathrm{B_s^0} \to \mu^+\mu^-$ avec une signifiance supérieure à 3σ (3.4σ) est obtenue pour la première fois. Pour $\mathrm{B^0} \to \mu^+\mu^-$ les données observées demeurent compatibles avec le bruit de fond mais permettent d'exclure avec un niveau de confiance de 95% la présence d'un signal avec un rapport d'embranchement de plus de 9.4×10^{-10}. Etant donnée la signifiance du signal $\mathrm{B_s^0} \to \mu^+\mu^-$, le rapport d'embranchement peut être mesuré dans les données. Pour cela, une fonction de densité de probabilité des distributions en masse des différentes contributions de bruit de fond et de signal, est ajustée simultanément aux distributions observées dans chaque catégorie de discriminant multivarié en minimisant la fonction de vraisemblance. Les rapports d'embranchement des deux signaux sont des paramètres libres de l'ajustement et le minimum de vraisemblance est obtenu pour :

$$\mathcal{B}(\mathrm{B_s^0} \to \mu^+\mu^-) \;=\; (3.2^{+1.4}_{-1.2}\ \text{stat.}\ ^{+0.5}_{-0.3}\ \text{syst.}) \times 10^{-9}\ \text{at}\ 68\%\ \text{C.L.}. \qquad (6)$$

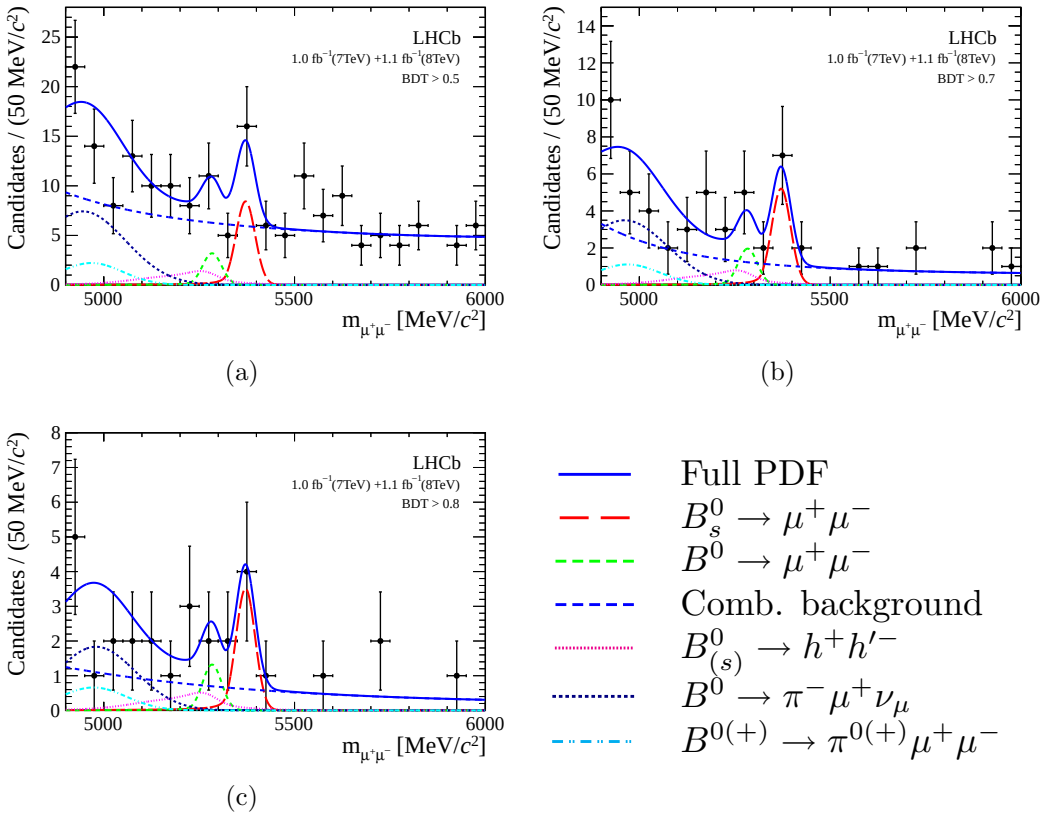


Figure 8: Distribution de masse invariante pour les candidats des données 2012 ($1.0\,\text{fb}^{-1}$) et 2011 ($1.1\,\text{fb}^{-1}$) ayant un BDT supérieur à 0.5 (a), 0.7 (b), and 0.8 (c).

La distribution en masse des catégories avec la plus grande sensibilité au signal est montrée dans la Figure 8.

Les collaboration ATLAS et CMS ont aussi apporté des résultats intéressants sur la recherche de $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ et $\mathsf{B}^0 \to \mu^+\mu^-$ [15, 16]. Avec une résolution en masse supérieure à 60 MeV, ATLAS est seulement sensible à la somme des deux signaux et en supposant que la contribution de $\mathsf{B}^0 \to \mu^+\mu^-$ est négligeable par rapport à celle de $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$, les données observées permettent d'exclure avec un niveau de confiance de 95 % tous les rapports d'embranchement de $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ supérieur à 22×10^{-9}. Au contraire CMS est sensible aux deux modes et les données observées permettent d'exclure avec un niveau de confiance de 95 % les rapports d'embranchement de $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-(\mathsf{B}^0 \to \mu^+\mu^-)$ supérieur à $7.7 \times 10^{-9}(18 \times 10^{-10})$.

Les résultats des trois collaborations peuvent être combinés avec la méthode CLs en prenant soin de traiter les corrélations entre les paramètres com-

muns (f_s/f_d par exemple). Cette combinaison permet d'obtenir une limite sur le rapport d'embranchement de $\mathsf{B}^0 \to \mu^+\mu^-$ de 8.1×10^{-10} et une signifiance pour le signal $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ de 3.6σ.

Interprétation des Résultats

Avec les premières mesures du rapport d'embranchement de $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$, des effets jusqu'alors négligés dans l'interprétation de ces résultats ont été considérés. Tout d'abord un effort a été fait pour accorder les définitions de la grandeur calculée théoriquement – le rapport d'embranchement à $t = 0$ – et de celle mesurée expérimentalement – le rapport d'embranchement intégré dans le temps. Pour les désintégrations de mésons B^0_s ces deux grandeurs sont différentes et sont reliées par :

$$\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-) \;=\; \frac{1 + y_s \mathcal{A}_{\Delta\Gamma}}{1 - y_s^2} \times \mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-)_{t=0}, \tag{7}$$

avec

$$y_s \;=\; \frac{\Gamma_\mathsf{L} - \Gamma_\mathsf{H}}{\Gamma_\mathsf{L} + \Gamma_\mathsf{H}} \tag{8}$$

et

$$\mathcal{A}_{\Delta\Gamma} \;=\; \frac{\Gamma_{\mathsf{B}^0_{\mathsf{s,H}} \to \ell^+\ell^-} - \Gamma_{\mathsf{B}^0_{\mathsf{s,L}} \to \ell^+\ell^-}}{\Gamma_{\mathsf{B}^0_{\mathsf{s,H}} \to \ell^+\ell^-} + \Gamma_{\mathsf{B}^0_{\mathsf{s,L}} \to \ell^+\ell^-}}. \tag{9}$$

Cette relation traduit l'effet du mélange entre le B^0_s et le $\overline{\mathsf{B}}^0_\mathsf{s}$ et de la différence de largeur des états propres de masse du B^0_s. Cependant l'acceptante en temps de l'analyse n'étant pas triviale, les résultats dépendent de l'hypothèse faite sur la valeur de $\mathcal{A}_{\Delta\Gamma}$ comme le montre la Figure 9. Loin d'être un facteur limitant, cette dépendance, une fois prise en compte, permet de contrainte plus fortement l'espace des phases ouvert pour la nouvelle physique. Par exemple dans les modèles ou seules des contributions scalaires sont autorisées, la prise en compte de la dépendance des résultats avec $\mathcal{A}_{\Delta\Gamma}$ apporte une réduction supplémentaire de l'espace des phases de 30 % comme le montre la Figure 10. Cette étude montre également que $\mathcal{A}_{\Delta\Gamma}$ est une observable intéressante bien qu'elle soit difficilement accessible expérimentalement.

Recontruction Topologique Inclusive

La mesure d'autres modes comme $\mathsf{B}^0_\mathsf{s} \to \tau^+\tau^-$ est aussi intéressante pour tester tous les modèles avec des couplages à la troisième génération spéciaux.

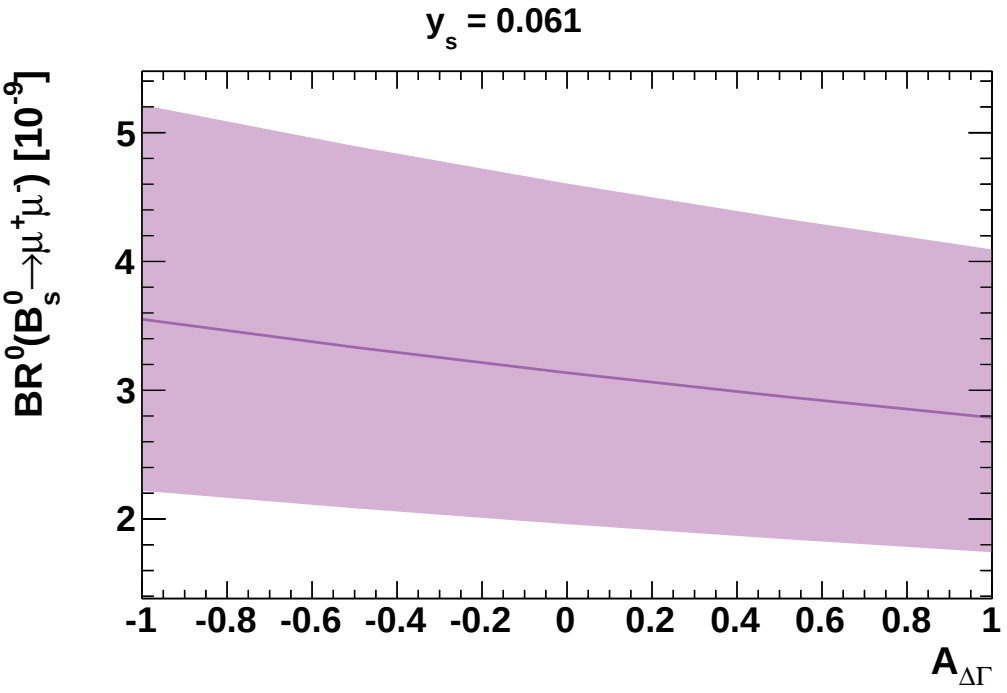
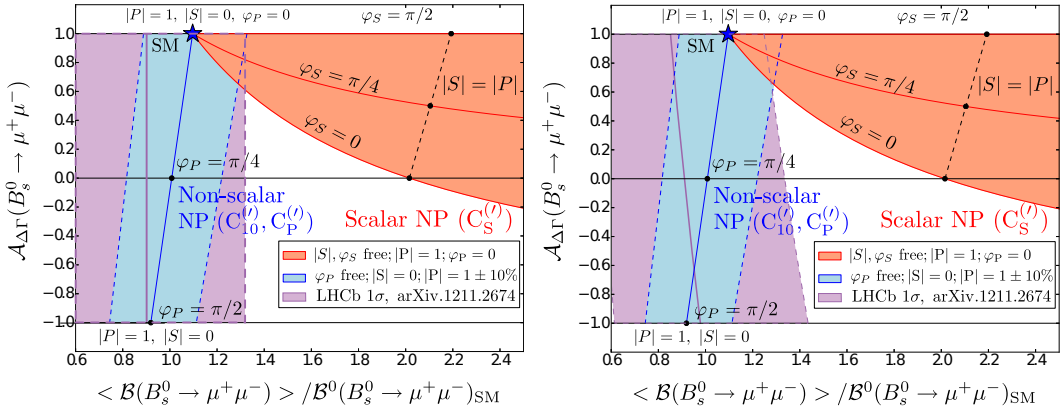


Figure 9: $\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-)$ mesuré en fonction of $\mathcal{A}_{\Delta\Gamma}$.

Cependant ce canal est expérimentalement beaucoup plus délicat. En effet les leptons τ ont un temps de vie de $c\tau = 87.11\,\mu$m et doivent donc être reconstruits à partir de leurs produits de désintégration qui contiennent nécessairement un ou plusieurs neutrinos. Le mode le plus simple et abondant permettant de reconstruire les deux τ est $\tau^- \to \pi^-\pi^-\pi^+\nu_\tau$. Le rapport d'embranchement effectif de la désintégration recherchée est alors de $(6.55 \pm 0.54) \times 10^{-9}$ avec un état final à six particules. Afin de reconstruire un tel état final, une méthode originale a été mise en place, inspirée par l'algorithme ZVTOP [18] développé à SLD. Cette méthode permet d'effectuer une reconstruction inclusive de l'évènement à partir de sa topologie. Une fonction de vertex est construire qui donne à chaque point de l'espace la probabilité d'avoir un vertex. Les vertex s'obtiennent alors en cherchant les maximums de cette fonction et en leur assignant un ensemble de traces. La Figure 11 montre une projection de cette fonction pour un évènement $\mathsf{B}^0_\mathsf{s} \to \tau^+\tau^-$ où les deux maximums, correspondants au vertex des deux τ apparaissent clairement. Cette méthode peut être intéressante pour reconstruire des modes, comme $\mathsf{B}^0_\mathsf{s} \to \tau^+(\pi^+\pi^+\pi^-\bar{\nu}_\tau)\tau^-(\pi^-\pi^-\pi^+\nu_\tau)$, où les traces de l'état final ne présentent pas de particularités cinématiques nota-



(a) (b)

Figure 10: Régions permises dans le plan $\mathcal{B}(\mathsf{B}_\mathsf{s}^0\to\mu^+\mu^-)/\mathcal{B}(\mathsf{B}_\mathsf{s}^0\to\mu^+\mu^-)_{\rm SM}\times\mathcal{A}_{\Delta\Gamma}$ si les contributions de la nouvelle physique sont scalaires (rouge) ou non-scalaire (bleue) sur imprimée avec la contrainte expérimentale (violette) avec (b) ou sans (a) la dépendance en $\mathcal{A}_{\Delta\Gamma}$ [17].

bles mais leur configuration par rapport au reste des traces de l'évènement (isolation par exemple) le peuvent.

Comparée a une reconstruction combinatoire classique, cette nouvelle méthode permettrait d'augmenter l'efficacité de reconstruction de $\mathsf{B}_\mathsf{s}^0\to\tau^+\tau^-$ de 21 %. Ce résultat est encourageant appelle des études plus approfondies pour évaluer la faisabilité d'une analyse basée sur cette méthode de reconstruction. Il a aussi été montré que la fonction de vertex peut être utilisée dans d'autres analyses comme par exemple $\mathsf{B}_\mathsf{s}^0\to\mu^+\mu^-$ où elle permet de construire de nouvelles variables liées à l'isolation des muons qui présentent une bonne séparation du signal et du bruit de fond combinatoire sans être corrélées aux variables déjà utilisées.

Conclusion

En conclusion, la première évidence de la désintégration $\mathsf{B}_\mathsf{s}^0\to\mu^+\mu^-$ marque une étape majeure dans la recherche de nouvelle physique. Les limites de plus en plus contraignante sur les rapports d'embranchement de $\mathsf{B}_{(\mathsf{s})}^0\to\mu^+\mu^-$ ainsi que les premières mesures de celui de $\mathsf{B}_\mathsf{s}^0\to\mu^+\mu^-$ apportées par LHCb durant ces trois dernières années [19, 20, 21, 22, 23] ont réduit énormément l'espace

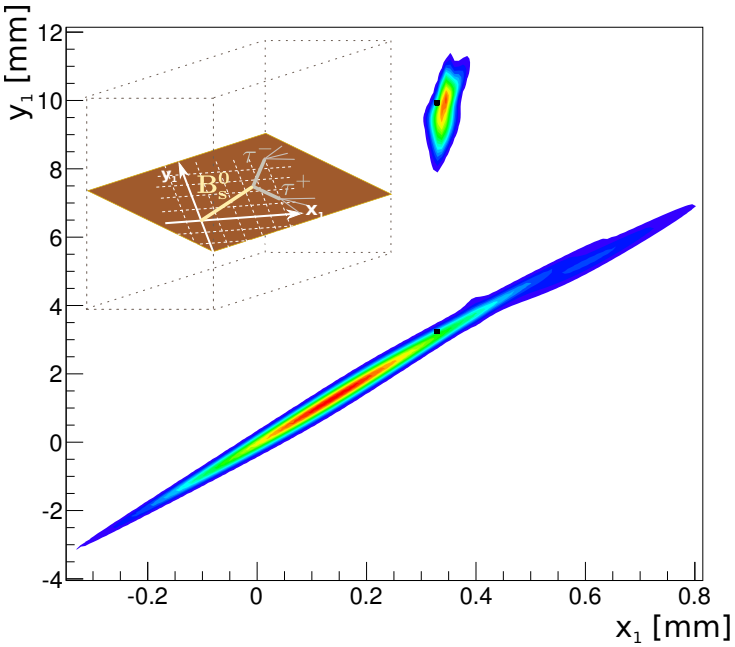


Figure 11: Fonction de vertex obtenue avec les six pions issus de la désintégration d'un événement $B_s^0 \to \tau^+\tau^-$ simulé. La fonction est projetée dans le plan de désintégration. L'origine de repère est définie à la vraie position du vertex du B_s^0et l'axe y passe par la vraie position des deux vertex de désintégration des τ (points noirs).

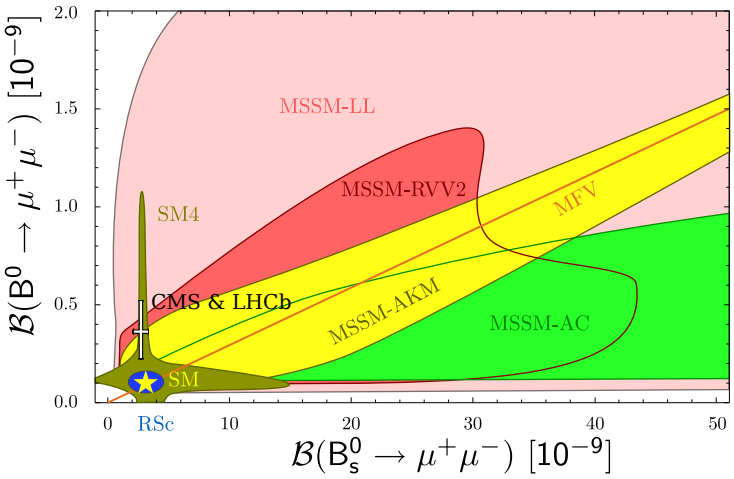


Figure 12: Prédictions pour les rapports d'embranchement de $B_s^0 \to \mu^+\mu^-$ et $B^0 \to \mu^+\mu^-$ dans les modèles à violation minimale de la saveur (MFV), le MS avec une quatrième génération (SM4), le modèle Randall-Sundrum (RSc) avec protection custodiale [4], quatre modèles minimaux de supersymétrie (MSSM), Agashe and Carone (AC) [5], Ross, Velasco-Sevilla and Vives (RVV2) [6], Antusch, King and Malinsky (AKM10) [7] et un modèle avec des courants gauches seulement (LL) [8]. La prédiction du MS est représentée par l'étoile et la moyenne des résultats de CMS et de LHCb [24] par la croix blanche (modifié de [9]).

de phase ouvert pour la nouvelle physique comme le montre la Figure 12 où est reporte les derniers résultat combinés des collaborations CMS et LHCb publiés pendant l'écriture de cette thèse [24].

La compatibilité des résultats avec le MS implique que cette analyse est vouée à devenir une mesure de précision. Dans cette optique, la prise en compte de la dépendance des résultats avec $\mathcal{A}_{\Delta\Gamma}$ est extrêmement importante. D'ores et déjà, sa prise en compte permet, dans certains modèles, d'obtenir une réduction supplémentaire de l'espace des phases de la NP de 30 %. Enfin le travail d'exploration de la nouvelle physique doit être poursuivit dans d'autre modes de désintégrations comme $B_s^0 \to \tau^+\tau^-$. La méthode de reconstruction topologique inclusive développée pour ce canal pourrait permettre d'augmenter l'efficacité de reconstruction de 21 % par rapport à une reconstruction combinatoire standard. Cette méthode peut également s'avérer extrêmement puissante pour étudier d'autres canaux comme il a déjà été montré par exemple pour $B_s^0 \to \mu^+\mu^-$.

Table 1: Nombres attendus d'évènements de bruit de fond combinatoire, bruit de fond $B^0_{(s)} \to h^+h'^-$ identifié incorrectement et de signal MS comparés aux nombres d'évènements observés dans les fenêtres de masse de $B^0_s \to \mu^+\mu^-$ et de $B^0 \to \mu^+\mu^-$ dans chaque catégorie de BDT pour les données 2011.

Mode	BDT bin	0.0 – 0.25	0.25 – 0.4	0.4 – 0.5	0.5 – 0.6	0.6 – 0.7	0.7 – 0.8	0.8 – 0.9	0.9 – 1
$B^0_s \to \mu^+\mu^-$	Fond Comb.	1880^{+33}_{-33}	$55.5^{+3.0}_{-2.9}$	$12.1^{+1.4}_{-1.3}$	$4.16^{+0.88}_{-0.79}$	$1.81^{+0.62}_{-0.51}$	$0.77^{+0.52}_{-0.38}$	$0.47^{+0.48}_{-0.36}$	$0.24^{+0.44}_{-0.20}$
	MisId $B^0_{(s)} \to h^+h'^-$	$0.129^{+0.066}_{-0.050}$	$0.066^{+0.024}_{-0.019}$	$0.052^{+0.018}_{-0.015}$	$0.047^{+0.015}_{-0.013}$	$0.053^{+0.017}_{-0.014}$	$0.050^{+0.016}_{-0.013}$	$0.052^{+0.017}_{-0.014}$	$0.049^{+0.018}_{-0.014}$
	Signal MS	$2.70^{+0.81}_{-0.80}$	$1.30^{+0.27}_{-0.23}$	$1.03^{+0.20}_{-0.17}$	$0.92^{+0.15}_{-0.13}$	$1.06^{+0.17}_{-0.15}$	$1.10^{+0.17}_{-0.15}$	$1.26^{+0.20}_{-0.17}$	$1.31^{+0.28}_{-0.25}$
	Observation	1818	39	12	6	1	2	1	1
$B^0 \to \mu^+\mu^-$	Fond Comb.	1995^{+34}_{-34}	$59.2^{+3.3}_{-3.2}$	$12.6^{+1.6}_{-1.5}$	$4.44^{+0.99}_{-0.86}$	$1.67^{+0.66}_{-0.54}$	$0.75^{+0.58}_{-0.40}$	$0.44^{+0.57}_{-0.38}$	$0.22^{+0.48}_{-0.20}$
	MisId $B^0_{(s)} \to h^+h'^-$	$0.78^{+0.38}_{-0.29}$	$0.40^{+0.14}_{-0.10}$	$0.311^{+0.107}_{-0.079}$	$0.280^{+0.092}_{-0.068}$	$0.314^{+0.103}_{-0.076}$	$0.297^{+0.096}_{-0.071}$	$0.309^{+0.101}_{-0.075}$	$0.296^{+0.107}_{-0.079}$
	$B^0_s \to \mu^+\mu^-$ SM	$0.43^{+0.13}_{-0.13}$	$0.205^{+0.044}_{-0.037}$	$0.163^{+0.032}_{-0.027}$	$0.145^{+0.025}_{-0.021}$	$0.168^{+0.029}_{-0.025}$	$0.174^{+0.029}_{-0.024}$	$0.199^{+0.033}_{-0.029}$	$0.206^{+0.046}_{-0.040}$
	Signal MS	$0.328^{+0.096}_{-0.097}$	$0.158^{+0.030}_{-0.027}$	$0.125^{+0.022}_{-0.019}$	$0.112^{+0.016}_{-0.015}$	$0.129^{+0.019}_{-0.017}$	$0.134^{+0.018}_{-0.016}$	$0.153^{+0.022}_{-0.019}$	$0.159^{+0.032}_{-0.029}$
	Observation	1904	50	20	5	2	1	4	1

Table 2: Nombres attendus d'évènements de bruit de fond combinatoire, bruit de fond $B^0_{(s)} \to h^+h'^-$ identifié incorrectement, et de signal MS comparés aux nombres d'évènements observés dans les fenêtres de masse de $B^0_s \to \mu^+\mu^-$ et de $B^0 \to \mu^+\mu^-$ dans chaque catégorie de BDT pour les données 2012.

Mode	BDT bin	0.0 – 0.25	0.25 – 0.4	0.4 – 0.5	0.5 – 0.6	0.6 – 0.7	0.7 – 0.8	0.8 – 1.0
$B^0_s \to \mu^+\mu^-$	Fond Comb.	2345^{+40}_{-40}	$56.7^{+3.0}_{-2.9}$	$13.1^{+1.5}_{-1.4}$	$4.42^{+0.91}_{-0.81}$	$2.10^{+0.67}_{-0.56}$	$0.35^{+0.42}_{-0.22}$	$0.39^{+0.33}_{-0.21}$
	MisId $B^0_{(s)} \to h^+h'^-$	$0.250^{+0.083}_{-0.068}$	$0.145^{+0.049}_{-0.040}$	$0.081^{+0.027}_{-0.023}$	$0.075^{+0.024}_{-0.020}$	$0.071^{+0.023}_{-0.019}$	$0.062^{+0.021}_{-0.017}$	$0.104^{+0.034}_{-0.028}$
	Signal MS	$3.69^{+0.59}_{-0.52}$	$2.14^{+0.37}_{-0.33}$	$1.20^{+0.21}_{-0.18}$	$1.16^{+0.18}_{-0.16}$	$1.17^{+0.18}_{-0.16}$	$1.15^{+0.19}_{-0.17}$	$2.13^{+0.33}_{-0.29}$
	Observation	2274	65	19	5	3	1	3
$B^0 \to \mu^+\mu^-$	Fond Comb.	2491^{+42}_{-42}	$59.5^{+3.3}_{-3.2}$	$13.9^{+1.6}_{-1.5}$	$4.74^{+1.00}_{-0.89}$	$2.10^{+0.74}_{-0.61}$	$0.55^{+0.50}_{-0.31}$	$0.29^{+0.34}_{-0.19}$
	MisId $B^0_{(s)} \to h^+h'^-$	$1.49^{+0.50}_{-0.36}$	$0.86^{+0.29}_{-0.22}$	$0.48^{+0.16}_{-0.12}$	$0.44^{+0.15}_{-0.11}$	$0.42^{+0.14}_{-0.10}$	$0.369^{+0.126}_{-0.093}$	$0.62^{+0.21}_{-0.15}$
	$B^0_s \to \mu^+\mu^-$ SM	$0.627^{+0.104}_{-0.091}$	$0.363^{+0.066}_{-0.057}$	$0.204^{+0.036}_{-0.032}$	$0.197^{+0.032}_{-0.027}$	$0.199^{+0.032}_{-0.028}$	$0.196^{+0.034}_{-0.030}$	$0.362^{+0.058}_{-0.051}$
	Signal MS	$0.442^{+0.062}_{-0.057}$	$0.256^{+0.040}_{-0.036}$	$0.144^{+0.022}_{-0.020}$	$0.139^{+0.019}_{-0.017}$	$0.140^{+0.019}_{-0.018}$	$0.138^{+0.021}_{-0.019}$	$0.255^{+0.035}_{-0.031}$
	Observation	2433	59	19	3	2	2	2

Bibliography

[1] J. Butler et al. *Report of the Quark Flavor Physics Working Group.* ArXiv:1311.1076, (2013).

[2] R. Aleksan, et al. *Physics Briefing Book: Input for the Strategy Group to draft the update of the European Strategy for Particle Physics.* CERN-ESG-005 (2013). Open Symposium held in Cracow from 10th to 12th of September 2012.

[3] R. Giles et al. *Two-Body Decays of* B *Mesons.* Physical Review, **D30** (1984), 2279.

[4] M. Blanke, et al. *Rare* K *and* B *Decays in a Warped Extra Dimension with Custodial Protection.* Journal of High Energy Physics, **0903** (2009), 108.

[5] K. Agashe et al. *Supersymmetric flavor models and the* $\mathsf{B} \to \phi \mathsf{K}^0_S$ *anomaly.* Physical Review, **D68** (2003), 035017.

[6] G. G. Ross, et al. *Spontaneous CP violation and nonAbelian family symmetry in SUSY.* Nuclear Physics, **B692** (2004), 50–82.

[7] S. Antusch, et al. *Solving the SUSY Flavour and CP Problems with SU(3) Family Symmetry.* Journal of High Energy Physics, **0806** (2008), 068.

[8] L. J. Hall et al. *A Geometry of the generations.* Physical Review Letter, **75** (1995), 3985–3988.

[9] D. M. Straub. *Overview of Constraints on New Physics in Rare B Decays.* ArXiv:1205.6094, (2012).

[10] T. Sjostrand, et al. *PYTHIA 6.4 Physics and Manual.* Journal of High Energy Physics, **0605** (2006), 026.

[11] T. Sjostrand, et al. *A Brief Introduction to PYTHIA 8.1.* Computer Physics Communications, **178** (2008), 852–867.

[12] LHCb Collaboration. *Measurement of b-hadron production fractions in* 7 TeVpp *collisions.* Physical Review, **D85** (2012), 032008.

[13] LHCb Collaboration. *Measurement of the fragmentation fraction ratio* f_s/f_d *and its dependence on* B *meson kinematics.* Journal of High Energy Physics, **1304** (2013), 001.

[14] A. L. Read. *Presentation of search results: The CL_s technique.* Journal of Physics, **G28** (2002), 2693–2704.

[15] G. Aad et al. *Search for the decay* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *with the ATLAS detector.* Physics Letter, **B713** (2012), 387–407.

[16] CMS Collaboration. *Search for* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *and* $\mathsf{B}^0 \to \mu^+\mu^-$ *decays in pp collisions at 7 TeV.* Physical Review Letter, **107** (2011), 191802.

[17] K. D. Bruyn, et al. *Probing New Physics via the* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *Effective Lifetime.* Physical Review Letter, **109** (2012), 041801.

[18] D. Jackson. *A topological vertex reconstruction algorithm for hadronic jets.* Nuclear Instruments and Methods in Physics Research, **A388** (1997) 1-2, 247–253.

[19] LHCb Collaboration. *Search for the rare decays* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *and* $\mathsf{B}^0 \to \mu^+\mu^-$. Physics Letter, **B699** (2011), 330–340.

[20] LHCb Collaboration. *Search for the rare decays* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *and* $\mathsf{B}^0 \to \mu^+\mu^-$. Physics Letter, **B708** (2012), 55–67.

[21] LHCb Collaboration. *Strong constraints on the rare decays* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *and* $\mathsf{B}^0 \to \mu^+\mu^-$. Physical Review Letter, **108** (2012), 231801.

[22] LHCb Collaboration. *First evidence for the decay* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$. Physical Review Letter, **110** (2013), 021801.

[23] LHCb Collaboration. *Measurement of the* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *branching fraction and search for* $\mathsf{B}^0 \to \mu^+\mu^-$ *decays at the LHCb experiment.* Physical Review Letter, **111** (2013) 10, 101805.

[24] CMS and LHCb Collaborations. *Combination of results on the rare decays* $\mathsf{B}_{(\mathsf{s})}^0 \to \mu^+\mu^-$ *from the CMS and LHCb experiments.* LHCb-CONF-2013-012 (2013).

Contents

Introduction **29**

1 Probing High Energy Physics with $B^0_{(s)} \to \ell^+\ell^-$ **31**
1.1 The Standard Model of Particle Physics 31
1.1.1 Quantum Field Theory 32
1.1.2 The Standard Model Fields 34
1.1.3 Electroweak Unification 38
1.1.4 The Higgs Mechanism Idea 38
1.1.5 Fermions Masses and CKM Matrix 39
1.1.6 The Standard Model Shortcomings 41
1.2 Effective Field Theory 42
1.2.1 Different Energy Scales Separation 42
1.2.2 Hadron Weak Decay Effective Field Theory 43
1.3 $B^0_{(s)} \to \ell^+\ell^-$ Observables 44
1.3.1 $B^0_{(s)}$-mixing and $\mathcal{A}_{\Delta\Gamma}$ 45
1.3.2 Branching Fraction Expression 46
1.4 Standard Model Predictions 48
1.4.1 Standard Model Contribution Evaluation 48
1.4.2 $B^0_{(s)} \to \mu^+\mu^-$ SM Predictions 50
1.4.3 $B^0_{(s)} \to \tau^+\tau^-$ SM Predictions 51
1.5 Phenomenology of $B^0_s \to \ell^+\ell^-$ Beyond the Standard Model . . 51
1.5.1 Model Independent Discussion 52
1.5.2 Model Dependent Discussion 55
1.6 Conclusion 63
Bibliography 65

2 The LHCb Experiment **71**
2.1 The LHC Environment 71
2.1.1 The CERN Accelerator Complex 71
2.1.2 B Physics at the LHC Environment 72

2.2 The LHCb Experiment 79
2.2.1 General Description 79
2.2.2 Tracking 79
2.2.3 Particle Identification 88
2.2.4 Trigger 93
2.3 Conclusion 98
Bibliography 99

3 The First Evidence of $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ 101
3.1 Overview of the Analysis 103
3.1.1 Introduction 103
3.1.2 Simulated Data Sample 105
3.1.3 Trigger 106
3.1.4 Initial Selection 108
3.1.5 Tight Selection and BDT 109
3.1.6 Classification 115
3.1.7 Classifier Calibration 116
3.1.8 Normalisation 120
3.1.9 Limits and $\mathrm{CL_s}$ Method 124
3.2 Binning Optimisation 128
3.2.1 Quantifying the Sensitivity 128
3.2.2 Evaluation of the Expectation Signal and Background Bin Content 130
3.2.3 Results 133
3.2.4 Discussion of the Performances 136
3.3 Upper Limits and p-value Extraction 137
3.3.1 Implementation of the $\mathrm{CL_s}$ Method 139
3.3.2 Pseudo-experiment Generation 140
3.3.3 Test Statistic Computation 143
3.3.4 Sensitivity Projection 144
3.4 First Evidence of $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ 151
3.4.1 Expectations 151
3.4.2 Observations 155
3.4.3 Extraction of $\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-)$ 158
3.5 Combining with other LHC experiments 162
3.5.1 ATLAS Analysis 166
3.5.2 CMS Analysis 167
3.5.3 Results and Conclusions 167
3.6 Conclusion 174
Bibliography 176

4 Interpretation of the Results **181**
4.1 Model Dependent Results . . . 181
4.1.1 B_s^0-$\overline{B}_s^0$ Mixing Effect . . . 182
4.1.2 Model Dependent Normalisation . . . 182
4.1.3 New Definition of the Normalisation Factor . . . 183
4.1.4 Model Dependent BDT PDF . . . 184
4.1.5 Corrected Results . . . 187
4.2 Theoretical Predictions Confronting Experimental Results . . . 189
4.2.1 Model Independent Constraints . . . 190
4.2.2 Model Dependent Constraints . . . 193
4.3 Conclusion . . . 195
Bibliography . . . 196

5 An Inclusive Event Vertex Reconstruction **199**
5.1 Motivations . . . 199
5.2 General Algorithm . . . 200
5.2.1 Vertex Function . . . 200
5.2.2 Algorithm . . . 201
5.2.3 Algorithm Parameters . . . 203
5.3 Maximum Finder Algorithms . . . 203
5.3.1 LHCb Environment . . . 203
5.3.2 Parabolic Extrapolation . . . 204
5.3.3 Step Direction Evaluation . . . 206
5.4 $B_{(s)}^0 \to \tau^+\tau^-$ Reconstruction . . . 208
5.4.1 Signal Description . . . 208
5.4.2 Signal Candidate Reconstruction . . . 209
5.5 Isolation Tool . . . 211
5.5.1 $B_{(s)}^0 \to \mu^+\mu^-$ Combinatorial Background . . . 211
5.5.2 Vertex Function Based Isolation Variables . . . 211
5.6 Conclusion . . . 214
Bibliography . . . 222

Conclusions and Perspectives **223**

Appendices

A Variables Definitions **229**
A.1 Track Isolation . . . 229
A.2 B Isolation or "CDF Isolation" . . . 230

B Latest Analyses Update **231**
B.1 LHCb Analysis 231
B.2 CMS Analysis 234
B.3 CMS and LHCb Analysis Combination 236
Bibliography 240

Introduction

The physics theory used to describe matter and its interactions at the microscopic scale, or equivalently at high energy, has converged, after a long standing history, to a paradigm called the Standard Model (SM) of particle physics. Paradoxically, this paradigm is to a large extent self-negating. Indeed, based on theoretical and experimental grounds, the particle physics community agrees today to conceive the SM only as an effective low energy version of a more general theory to be discovered. Theoretical and experimental searches have endeavoured for decades to find hints of this new high energy physics. One of the main portals towards new physics are the heavy flavour meson decays. After establishing the fundamental concepts at the basis of the SM and its extensions, Chapter 1 explains how the study of these decays can give access to new physics. In particular, it will show that $\mathsf{B}^0_\mathsf{s} \to \ell^+\ell^-$ processes are among the channels the most sensitive to physics beyond the SM.

Searches for these processes started thirty years ago at the CLEO experiment and are now pursued at the Large Hadron Collider (LHC). Chapter 2 describes the environment the LHC provides and explains how these operating conditions are exploited by the LHCb detector whose design and performances allow to reach among the highest sensitivities to these decays.

This remarkable sensitivity was demonstrated soon after the first LHC pp collisions and four consecutive analyses brought, each time, world best results on the $\mathsf{B}^0_{(\mathsf{s})} \to \mu^+\mu^-$ decays. Chapter 3 describes in details the fourth analysis which, with $2.1\,\text{fb}^{-1}$ of integrated luminosity, has obtained the first evidence of the $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ decay and a world best limit of the branching fraction of $\mathsf{B}^0 \to \mu^+\mu^-$. These results can be further improved by combining them with the results by two other LHC experiment, ATLAS and CMS.

The interpretations of these results are presented in Chapter 4. They translate in constraints on the phase space allowed for new physics. Based on considerations developed in Chapter 1 on a phenomenon called B-mixing, this chapter demonstrates that the experimental results are model dependent and that taking into account this model dependency allows to significantly

improve the sensitivity to new physics.

Finally, an original method to reconstruct vertices based on a global function returning at each space point a probability to have a vertex is developed in Chapter 5. This method is intended to be used to reconstruct $\mathsf{B}^0_{(\mathsf{s})} \to \tau^+\tau^-$ candidates and a feasibility study is presented. Searching for these modes is experimentally challenging as the τ need to be reconstructed from its decay products which unavoidably contain at least one neutrino. Thus, an inclusive event approach could be advantageous. However, the realm of application of this inclusive method is much wider than $\mathsf{B}^0_\mathsf{s} \to \tau^+\tau^-$. For instance it can be used to provide isolation variables to separate the $\mathsf{B}^0_{(\mathsf{s})} \to \mu^+\mu^-$ signal from its combinatorial background.

In summary, the works presented in this thesis all hinge around the searches for B meson decays to purely leptonic final states. The next section will explain why these searches are prominently important.

Chapter 1

Probing High Energy Physics with $\mathsf{B}^0_{(s)} \to \ell^+\ell^-$

The theory used to describe particle physics nowadays is the result of a long standing history, driven by both experimental and theoretical progresses. Despite these impressive progresses, this theory is not completely satisfactory and unsolved issues suggest that it is probably only an effective version at low energy of a more general theory. This chapter aims to explain why the $\mathsf{B}^0_{(s)} \to \ell^+\ell^-$ decays are among the best portals towards the new physics that this more general theory would imply.

The main concepts used in the current theory, called Standard Model (SM), are described in Section 1.1 then, on the basis of these concepts, an effective formulation of the theory is introduced in Section 1.2. From this formulation a general expression of the branching fraction is derived in Section 1.3. The SM and beyond SM phenomenologies are then described in Section 1.4 and Section 1.5. The latter section begins with a general discussion in terms of parameters (Wilson Coefficients) introduced in the effective formulation and is concluded by describing a few chosen SM extensions.

1.1 The Standard Model of Particle Physics

The Standard Model (SM) of Particle Physics is the model currently used to describe particle physics[1]. In its present formulation, it relies on Quantum Field Theories (QFT) to describe particles and their interactions. Section 1.1.1 sketches the general characteristics of quantum fields. The SM fields are described in Section 1.1.2, while Section 1.1.3 explains how these

[1]For historical perspectives on the building of the SM the reader is referred to [1].

fields can be merged under a more unified vision. This unification implies the so called Higgs mechanism which is exposed in Section 1.1.4. Section 1.1.5 makes the connection between this mechanism and the mixing between quarks of different flavours. Finally, despite the impressive success of the SM, we will see in Section 1.1.6 that some issues remain and lead to think about the SM as an effective version of a more general theory.

1.1.1 Quantum Field Theory

Quantum Field Theory provides the theoretical framework to describe elementary particles and their interactions in the SM. As a field theory it allows to describe systems with an undefined and non constant number of relativistic particles and as a quantum theory it catches the microscopic nature of these particles.

Real and Virtual Particles

Unlike classical fields where stable states can propagate with any energy, in a quantum field the stable propagating states have defined discrete energies [2, 3]. Intuitively these states can be interpreted as particles: the first allowed excited energy level of the field E_1 describes one particle of mass m_1 and momentum p_1 such that $E_1 = \sqrt{m_1^2 + p_1^2}$; then if the fields is excited up to its second allowed energy level E_2, another particle with mass m_2 and momentum p_2 arises such that $E_2 = \sqrt{m_1^2 + p_1^2} + \sqrt{m_2^2 + p_2^2}$. In this way the quantum field is able to describe a system with a non constant number of particles. The state corresponding to zero particle is called *ground state*.

In addition to these stable propagating states, the field can also sustain transient non stable propagating sates with any energy. However, these states cannot be interpreted as particles in the previous acceptation. Instead, they mediate the interactions between particles [4] and are called virtual particles. Since a virtual particle is not a stable propagating states, once it has been emitted, it must be reabsorbed by a time inversely proportional to its energy due to the uncertainty principle. This property will be used several times in the following discussions.

Feynman Diagram

Interactions between particles are usually represented with Feynman diagrams, like for example the electrons' and photons' scattering in Figure 1.1. In these diagrams, the propagating states are represented as external lines, and the energy and momentum exchange, mediated by virtual particles, are

represented as internal lines. Thus electrons and photons can be both real or virtual particles.

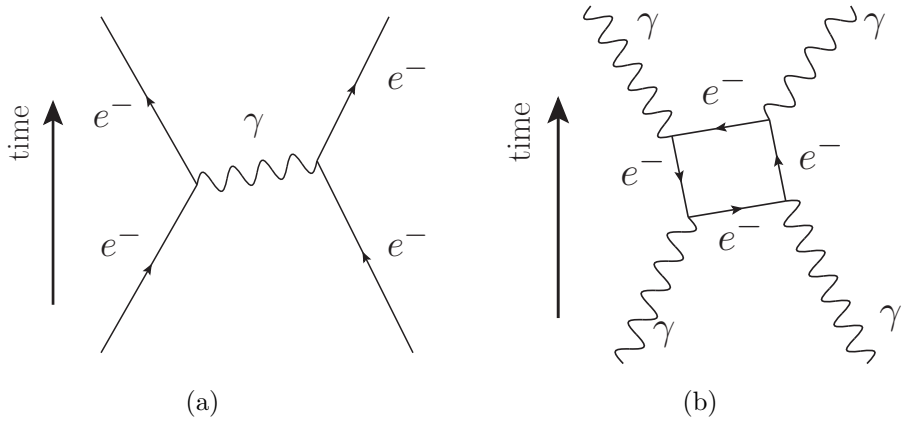


Figure 1.1: Dominant Feynman diagrams of electron (a) and photon (b) scattering.

External lines can be connected between each others by many other internal lines configurations as long as they respect some basic requirements[2]. Each of these configurations define a new Feynman diagram. The probability for the process to happen via a given Feynman diagram can be computed from the diagram itself using a set of Feynman rules. These rules stipulate in particular that for each vertex in the diagram the probability is multiplied by a constant proportional to the interaction strength squared. In quantum electro-dynamics (QED, the theory obtained by quantising electromagnetism), for example, this factor is the fine structure constant $\alpha \simeq 1/137$. Since α is smaller than 1.0, the more vertices in the diagrams, the smaller the diagram contributes to the transition. Hence, the series of all the possible Feynman diagrams is in fact a perturbative expansion of the transition amplitude in powers of α.

Renormalisation

The momentum and energy being conserved between the initial and final states, the kinematic properties of the virtual photon in Figure 1.1(a) are fully determined. However, for diagrams featuring internal loops, like in

[2]In quantum electro-dynamics (QED, the theory obtained by quantising electromagnetism) only three-prong vertices connecting one incoming and one outgoing electron or positron with a photon are allowed.

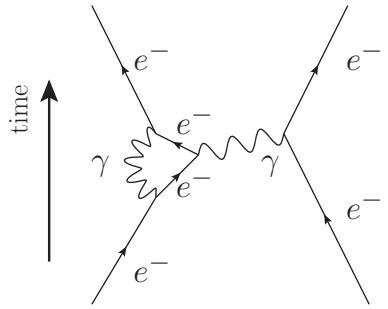


Figure 1.2: Feynman diagram contributing to the electrons scattering whose amplitude diverges.

Figure 1.2, the virtual particles momenta are no longer constrained by the external lines. Hence an infinite number of quantum states between the initial and final state are possible. As a result, the computation of the amplitude of these loop diagrams diverges. This divergence is an artefact which can be solved by renormalising the theory.

As explained in the previous section, once a virtual particle has been emitted, it must be reabsorbed by a time inversely proportional to its energy. Hence, in Figure 1.2 the size of the loop reduces as the energy circulating in it rises. At some point the loop is so small that it is equivalent to a point like vertex. The resulting diagram is then exactly the same as the one in Figure 1.1(a) which has already been accounted for in the expansion series. Hence, to avoid double counting, all the very energetic loop diagram configurations must be subtracted and the initial divergence cancels in this subtraction.

Technically, this subtraction is performed by redefining the couplings which technically implies to choose a *renormalisation scale* (i.e the scale at which the vertex in Figure 1.2 is said to be point like) and a *renormalisation scheme* which defines what the subtracted counter-terms contain and so how the couplings are redefined.

1.1.2 The Standard Model Fields

The SM fields are divided in two categories: the spin 1/2 fermions and the spin 1 bosons. These two field families are described briefly in the following and a classification is provided in Figure 1.3. These fields can interact via the strong, weak, and electromagnetic forces.

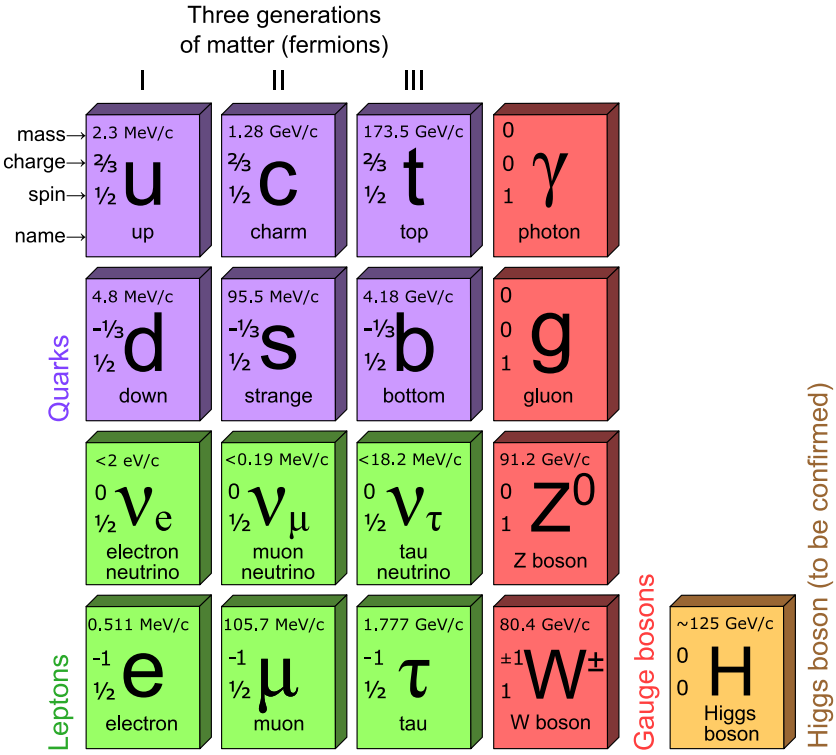


Figure 1.3: Field content of the Standard Model.

Fermions: Quarks and Leptons

The Standard Model contains three families of fermions[3]. Each family contains two leptons and two quarks[4] sensitive to different interactions.

Quarks are sensitive to the strong force, hence they carry a colour charge (red, green, blue and their associated anti-colour). They are also sensitive to the electromagnetic and weak interactions. Each family contains one up-type and one down-type quarks with electric charges equal to $+2/3$ and $-1/3$ of the absolute value of the electron electric charge.

These properties are replicated for each fermion family. Hence the only distinctive feature between families is the particle masses which increase with the generation number. The first family quarks are called *up* (u) and *down* (d), the second *charm* (c) and *strange* (s), and the third *beauty* (b also called *bottom*) and *top* (t).

As the strong force has infinite range and its strength does not diminish with increasing distance, an isolated colour charge would radiate an infinite energy. Hence quarks are, with the exception of the top quark, never found alone. Instead they group in pairs (called mesons) or triplets (called baryons) arranged to be globally colour-less. These phenomenon is called *hadronisation*[5]. In this thesis we study the decay of neutral B mesons which can be B^0_s or B^0. B^0_s contains a $\overline{\mathsf{b}}$ and a s quarks, and B^0 a $\overline{\mathsf{b}}$ and a d.

On the contrary *leptons* are not sensitive to the strong force. For each family there are two lepton species. The first species is sensitive to the electromagnetic and weak interaction, while the second interacts only weakly. For the first family the two species are the electron e^-, and the electronic neutrino ν_e. Their duplicate in the second family are the muon μ with its associated neutrino ν_μ and the tau τ with its neutrino ν_τ in the third. Likewise to the quarks the only difference between the three families is the mass of their particles. For the neutrinos, the masses are still unknown. The observations of neutrinos flavour oscillation [5] indicate that they cannot be zero.

The mass pattern and the replication of the other particles properties through families are still not understood. Hence it is interesting to study the decays of neutral B mesons to both muon and tau pairs as it could reveal some unexpected properties of this family replication.

[3]The SM does not explain the reason why only three families exist, which leads some theoreticians to postulate the existence of a fourth family (see Section 1.5).

[4]This peculiar symmetry is not explained in the SM and is the root of all the leptoquarks SM extensions (see Section 1.5).

[5]The top quark is too short-lived for the hadronisation mechanism to happen.

Gauge Boson

When a theory is mathematically formulated, all terms corresponding to non measurable quantities (for instance the absolute position), are linked to invariances of the theory under symmetries (translation in the previous example).

QFT formulation usually implies a Lagrangian density $\mathcal{L}$ containing the fields $\psi(x)$ and their derivatives $\partial\psi(x)$. In case of electromagnetism, the absolute phase of the the field is not measurable. This internal degrees of freedom can be arbitrarily chosen by multiplying the field by $e^{i\theta}$. The set of all these transformations is the *symmetry group* U(1) for which $\mathcal{L}$ is indeed invariant. However, there is no reason for this phase to be the same over the entire universe and for all times. Instead the phase could be fixed locally in a region with the dimension of our experiment and no reference to far-away distances. Hence the theory should be invariant under a continuous transformation:

$$\psi(x) \quad \rightarrow \quad e^{i\theta(x)}\psi(x). \tag{1.1}$$

Unfortunately under this continuous transformation the field derivative becomes:

$$\partial\psi(x) \quad \rightarrow \quad e^{i\theta(x)}\partial\psi(x) + ie^{i\theta(x)}\psi(x)\partial\theta(x), \tag{1.2}$$

and $\mathcal{L}$ is no longer invariant. To recover this invariance the derivative must be replaced by a covariant derivative which transforms like the field itself. This covariant derivative is obtained by introducing a new field called *gauge field*. With this field a more general invariant $\mathcal{L}$ can be written with a term describing gauge field kinematic energy. In QED the resulting gauge field is the photon, i.e. the electromagnetic force carrier. More generally, the gauge invariance principle allows to describe all SM forces with gauge bosons:

- as explained above, the QED gauge boson is the photon γ and arises by requiring invariance under the symmetry group $\mathrm{U}(1)_{EM}$.
- The weak interaction gauge bosons are the $\mathsf{W}^{\pm}$ and the Z^{0}. These bosons are massive and in the case of the $\mathsf{W}^{\pm}$ carry electric charge. They arise by requiring invariance of $\mathcal{L}$ under the symmetry group $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$.
- Finally, the QCD gauge bosons are the eight gluons. The gluons are massless and carry colour charges. They arise by requiring invariance under the symmetry group SU(3).

For a pedagogical introduction to gauge theory, the reader is referred to [6].

1.1.3 Electroweak Unification

Even though the weak and electromagnetic interactions appear to be very different processes, they can be viewed as a single force called electroweak interaction and described by requiring $\mathcal{L}$ to be invariant under the gauge group $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$. In this unified framework, the two matter fields, corresponding to the charged leptons and their neutrino, belong, in each family, to the same representation of the group. However, these two fields appear experimentally to be very different as the electron is massive while the neutrino is not[6]. In addition, the apparent disparity in the two interaction strengths suggests that the weak interaction is mediated by a massive boson[7] while the electromagnetic gauge boson, the photon, is massless. Thus, for the unification to be achieved despite these apparent differences, another process must be at stake to explain how the electroweak symmetry is broken. The favourite explanation to this symmetry breaking is the so called *Higgs mechanism*. The next section describes a simplified version of it.

1.1.4 The Higgs Mechanism Idea

Electroweak symmetry breaking in the SM requires a new spinless quantum field named after P. Higgs, one of the first theoreticians with F. Englert, R. Brout, G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble to introduce this idea in 1964 [7, 8, 9]. This field ϕ is in fact a complex scalar doublet and its two components are noted ϕ_1 and ϕ_2. This field is associated to a generic potential $V(\phi)$, defined with two parameters m and g:

$$V(\phi) \;=\; m^2(\phi_1^2 + \phi_2^2) + g(\phi_1^2 + \phi_2^2)^2, \tag{1.3}$$

and designed to be invariant when rotated on the axis perpendicular to ϕ_1 and ϕ_2.

This generic expression allows two types of shapes for the potential depending on the sign of m^2. Positive m^2 values describe paraboloid shapes and a negative m^2 the so-called Mexican hat shapes shown in Figure 1.4. Even though the potential can be transformed continuously from one topology to another by varying m^2, the theory behaves completely differently as soon as m^2 becomes negative.

When the potential has a paraboloid shape, the equilibrium is located at its bottom and the two fields ϕ_1 and ϕ_2 correspond to two degenerate particles with mass m. When the shape becomes a Mexican hat, the previous

[6] The evidence of neutrinos' masses [5] is ignored in this discussion.

[7] Due to the uncertainty principle an interaction mediated by a massive boson is short ranged, see Section 1.1.1.

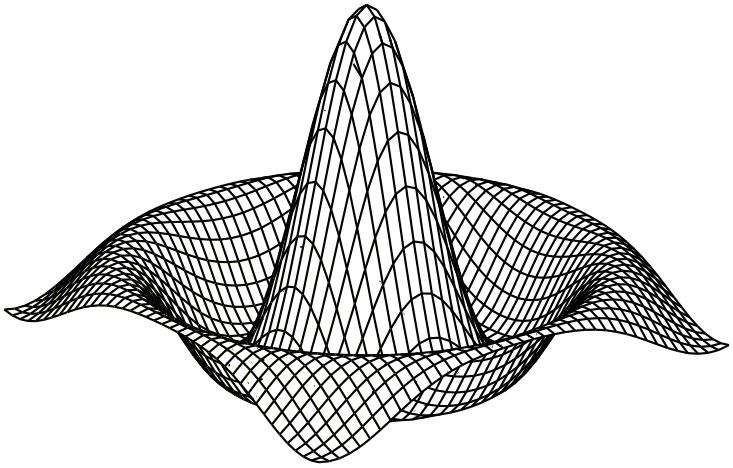


Figure 1.4: Higgs potential for $m^2 < 0$.

extremum is unstable and the field settles in a new minimum somewhere on the hat trough. At this position the two field components correspond to two different particles. A massive particle, called Higgs boson, corresponds to the radial vibration mode of the field and a massless particle, also called Goldstone boson, to the vibration mode along the trough.

The Goldstone boson responds to the force field in such a way that it produces a screening effect attenuating the force which consequently becomes short ranged. Since a short range force corresponds to massive gauge bosons, this screening effects produces in fact the mass of the bosons.

The m^2 term in Equation 1.3 evolved with the temperature of the universe. At some point in the universe evolution it went from positive to negative values which spontaneously produced the masses of the gauge bosons.

In July 2012, the ATLAS and CMS collaboration reported the observation of a new spinless particle [10, 11] which resembles as more the SM Higgs boson as its properties are measured [12, 13].

1.1.5 Fermions Masses and CKM Matrix

Fermion masses can also be produced by the Higgs mechanism by introducing in the theory the Yukawa interactions in the Lagrangian, $\mathcal{L}_Y$. In the case of quarks we will see that this interaction allows flavour changing interactions.

For the quarks, the Higgs field ϕ couples separately the right-handed up- and down-type singlet $u_{\mathrm{R},j}$ and $d_{\mathrm{R},j}$, to the left-handed doublet $\bar{Q}^{\mathrm{I}}_{\mathrm{L},i}$ according

to:

$$\mathcal{L}_Y \;=\; -Y^{\mathrm{d}}_{i,j}\bar{Q}_{\mathrm{L},i}\phi d_{\mathrm{R},j} - Y^{\mathrm{u}}_{i,j}\bar{Q}_{\mathrm{L},i}\epsilon\phi^* u_{\mathrm{R},j} + \mathrm{h.c.}, \tag{1.4}$$

where ϵ is the 2×2 antisymmetric tensor and i,j are generation labels. The Yukawa coupling constants $Y^{\mathrm{d,u}}$, are 3×3 non-diagonal complex matrices. They can be expressed as:

$$Y^{\mathrm{d,u}} \;=\; V_L^{\mathrm{d,u}} Y_D^{\mathrm{d,u}} V_R^{\mathrm{d,u}}, \tag{1.5}$$

where $Y_D^{\mathrm{d,u}}$ is a real diagonal matrix and $V_{L,R}^{\mathrm{d,u}}$ are hermitian matrices.

This expression allows to write $\mathcal{L}_Y$ with diagonal Yukawa couplings $Y_D^{\mathrm{d,u}}$ by redefining the fields. In particular, the left doublets are redefined as:

$$\tilde{Q}_{\mathrm{L}}^{\mathrm{d,u}} \;=\; \bar{Q}_{\mathrm{L}} V_L^{\mathrm{d,u}}. \tag{1.6}$$

However as Y^{d} and Y^{u} are different, the left doublet is redefined differently in the up and down part of Equation 1.4 and this allows tree-level flavour-changing charged interactions.

Indeed, since the weak interaction couples the left handed up and down type eigenstates, a term proportional to the non diagonal unitary matrix $V_L^{\mathrm{u}} V_L^{\mathrm{d}\dagger}$ arises in the coupling term and allows transition between flavours. This unitary matrix, $V_L^{\mathrm{u}} V_L^{\mathrm{d}\dagger}$, is actually the so-called Cabibbo-Kobayashi-Maskawa (CKM) matrix V_{CKM} [14, 15]:

$$V_{\mathrm{CKM}} \;=\; \begin{pmatrix} V_{\mathrm{ud}} & V_{\mathrm{us}} & V_{\mathrm{ub}} \\ V_{\mathrm{cd}} & V_{\mathrm{cs}} & V_{\mathrm{cb}} \\ V_{\mathrm{td}} & V_{\mathrm{ts}} & V_{\mathrm{tb}} \end{pmatrix}. \tag{1.7}$$

On the contrary, the Glashow-Iliopoulos-Maiani (GIM) mechanism [16] [8] forbids flavour changing neutral currents at tree levels. For a more formal introduction of the CKM matrix the reader is referred to Section 11 of the non abridged version of [18].

Wolfenstein proposed a parametrisation of the CKM matrix [19] which exhibits directly the hierarchy between the strengths of the different flavour transitions:

$$V_{\mathrm{CKM}} \;=\; \begin{pmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \tag{1.8}$$

[8] For a pedagogical description of the GIM mechanism the reader is referred to [17].

where the measured values [18] for λ, $\bar{\rho}$, A, and $\bar{\eta}$ are:

$$\begin{aligned} \lambda &= 0.22535 \pm 0.00065, & A &= 0.811^{+0.022}_{-0.012}, \\ \bar{\rho} &= 0.131^{+0.026}_{-0.013}, & \bar{\eta} &= 0.345^{+0.013}_{-0.014}. \end{aligned} \tag{1.9}$$

In the case of leptons, the SM contains no mixing terms as the neutrino are only left-handed. However observations of neutrino oscillation [5] suggest that an equivalent to the CKM matrix exist for the leptons. This matrix is usually called the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [20, 21].

1.1.6 The Standard Model Shortcomings

Over the past 40 years the SM has been very successful experimentally. However it cannot account for some observations and many theoretical aspects remain unsatisfactory. The SM is therefore believed to be an effective version of a more general theory. If this is the case, there must be some new physics beyond the SM.

In this section the main shortcomings of the SM are described. However, the candidates for physics beyond the SM will only be discussed in Section 1.5.2, as the framework used to describe the phenomenology they imply for the $\mathsf{B}^0_\mathsf{s} \to \ell^+\ell^-$ observables needs to be introduced earlier.

Experimental Issues

Measurements of neutrino flavour mixing [5] suggest that these particles have masses which is impossible to account for in the basic SM formulation.

In addition to this unaccounted phenomenon, many cosmological observations cannot be explained by the SM. The rotation of the galaxies indicates that there is more matter in the universe than what is visible [22, 23]. This interpretation is strengthened by the studies of fluctuations in the cosmic microwave background [24]. However, the SM provides no candidate for this invisible or dark matter which would in fact represent more than five times the mass of the standard matter in the universe.

Another compelling issue is the large asymmetry between matter and anti-matter which is observed nowadays, but cannot be explained in the SM if we assume an equal production of matter and anti-matter in the early universe. The only SM mechanism able to generate an asymmetry between matter and anti-matter is CP violation its magnitude in the SM is too small to account for the observed asymmetry.

Theoretical Issues

The SM has also some unsatisfactory theoretical aspects. The most obvious one is the very large number of free parameters (mainly coming from the Yukawa couplings). In addition, the SM framework is not suited to describe gravity. Finally the computation of the Higgs boson mass suffers from a quadratic divergence due to radiative corrections. In the SM this divergence can only be solved by fine tuning the bare mass of the Higgs boson so that a cancellation is operated with the radiative corrections. However this fine tuning is very unnatural.

1.2 Effective Field Theory

A given physical phenomenon usually involves physical processes acting at different distance scales and we are usually interested only by the phenomenology of a few of them. Section 1.2.1 shows how the different energy scale phenomenologies can naturally be separated. Technically, for hadron weak decays, this separation results in an effective Hamiltonian which allows to parametrise the decay amplitude with the so-called Wilson coefficients. This parametrisation is described in Section 1.2.2.

1.2.1 Different Energy Scales Separation

To explain how the separation of different energy scale phenomenologies arises naturally in particle physics, lets consider the β decay, $\pi^- \to e^- \bar{\nu}_e$. The simplest Feynman diagram contributing to this decay is the W^- exchange shown in Figure 1.5(a). The distance scale of this process is roughly the distance the W^- can propagate before being reabsorbed, i.e. $1/M_{\mathsf{W}^\pm}$ on the basis of the uncertainty principle. Lets now consider the diagram shown in Figure 1.5(b) which involves two processes acting at different distance scales: $1/M_{\mathsf{W}^\pm}$ for the W^+ exchange and $1/\Lambda_{QCD} \simeq 1/200\,\text{MeV}$ [18] for the gluon radiation. As $1/M_{\mathsf{W}^\pm}$ is much smaller than $1/\Lambda_{QCD}$, the W^+ exchange can be viewed as point-like interaction when the gluon exchange is studied, and the very detail of what happen in the vertex is irrelevant. Vice versa, at the energy scale of the W^+ exchange, the QCD asymptotic freedom makes the gluon exchange irrelevant. Hence the separation of short and long scale processes arises in a natural way.

This scale separation is crucial to probe new physics in the heavy flavour sector as new particles are sought in diagram loops which obviously imply short scale processes. In addition, the series of point-like vertices describing these short distance processes in the EFT are independent of the initial and

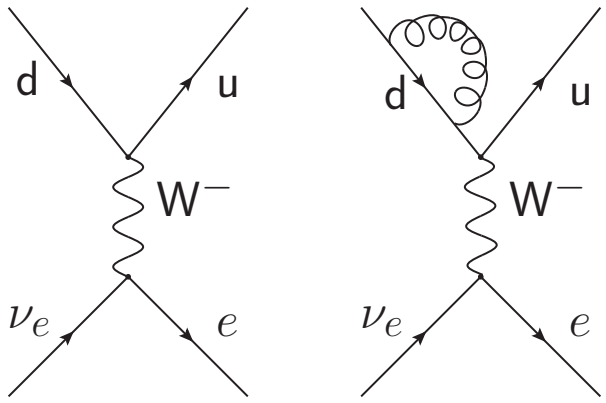


Figure 1.5: β decay Feynman diagrams with one (a) or two (b) distance scale(s).

final bound states and can therefore be tested in many different channels. An example of such global studies are provided in [25, 26, 27, 28, 29, 30, 31] for all the $\mathsf{b} \to \mathsf{s}$ and $\mathsf{b} \to \mathsf{s}\ell^+\ell^-$ transitions.

1.2.2 Hadron Weak Decay Effective Field Theory

In practice, the scale separation is performed with the operator product expansion (OPE) and an effective Hamiltonian is written for the hadron weak decays under the following generic structure [32]:

$$\mathcal{H}_{\mathsf{eff}} = \frac{G_F}{2} \sum_i V^i_{\mathrm{CKM}} C_i(\mu) Q_i(\mu). \tag{1.10}$$

The resulting $M \to F$ decay amplitudes are then expressed as:

$$\begin{aligned} \mathcal{A}(M \to F) &= \langle F|\mathcal{H}_{\mathsf{eff}}|M\rangle \\ &= \frac{G_F}{2} \sum_i V^i_{\mathrm{CKM}} C_i(\mu) \langle F|Q_i(\mu)|M\rangle. \end{aligned} \tag{1.11}$$

In Equations 1.10 and 1.11, G_F is the Fermi constant, and V^i_{CKM}, the CKM elements (see Section 1.1.5).

$C_i(\mu)$ are the so called Wilson coefficients [33]. They encode the physics of processes happening at an energy higher than μ, which involves usually contributions from $\mathsf{W}^\pm$, Z^0-bosons, top quark or new heavy particles if any.

Due to QCD asymptotic freedom, they can be calculated in perturbation theory as soon as μ is sufficiently large.

$Q_i(\mu)$ are local operators, they describe processes happening at an energy lower than μ. The resulting matrix elements $\langle F|Q_i(\mu)|M\rangle$ are computed with non-perturbative method, like lattice QCD, and are usually the dominant source of uncertainties in the amplitude evaluation.

The energy scale μ is arbitrary and usually set to the mass of the decaying hadron. As the choice of μ must not affect the final amplitude, a cancellation mechanism must happen in the EFT between the the operators and the Wilson coefficients μ-dependencies. In practice, if the energy scale is lowered, some high energy contributions in the Wilson coefficients are transferred into the low energy matrix elements so the amplitude no longer depends on μ. The matrix elements depend also on the renormalisation (see Section 1.1.1) scheme used in their computation. Hence, for the similar cancellation mechanism to work and the amplitude to be independent of the renormalisation scheme, the Wilson coefficients must absorb this dependency. In practice and in particular for the $\mathsf{B}^0_{(s)} \to \ell^+\ell^-$ processes (see Section 1.3.2), these two cancellation mechanisms have to be tested to ensure a reliable amplitude prediction.

The previous discussion already allows to better understand why the $\mathsf{B}^0_s \to \ell^+\ell^-$ processes are excellent probes of the SM and of the physics beyond it. First, these processes cannot be mediated at tree level in the SM[9], so the SM contribution to the Wilson coefficients is small. This contribution is all the more limited as these processes are helicity suppressed. Hence new physics, if any, could easily bring Wilson coefficients contributions as large as the SM one and therefore modify the decay amplitude up to an observable level. In addition, the final state being purely leptonic, the matrix elements can be evaluated with good accuracy, resulting in precise predictions for the SM and models beyond it. For more details on EFT and OPE, we report the reader to [34, 32, 35, 36].

1.3 $\mathsf{B}^0_{(s)} \to \ell^+\ell^-$ Observables

The decays of neutral $\mathsf{B}^0_{(s)}$ mesons offer a rich phenomenology with many observables [37]. This richness arises as the $\mathsf{B}^0_{(s)}$ undergoes a phenomenon called $\mathsf{B}^0_{(s)}$-mixing. Section 1.3.1 describes this phenomenon together with a related observable called $\mathcal{A}_{\Delta\Gamma}$. Then, on the basis on the effective field theory,

[9] Glashow, Iliopoulos, and Maiani [16] demonstrated that there is not tree level Flavour Changing Neutral Currents in the SM.

developed in the previous section, Section 1.3.2 introduces the expression of the $\mathsf{B}^0_\mathsf{s} \to \ell^+\ell^-$ branching fraction and $\mathcal{A}_{\Delta\Gamma}$.

1.3.1 $\mathsf{B}^0_{(\mathsf{s})}$-mixing and $\mathcal{A}_{\Delta\Gamma}$

This $\mathsf{B}^0_{(\mathsf{s})}$-mixing phenomenon happens as the neutral $\mathsf{B}^0_{(\mathsf{s})}$ mesons are produced in two flavour eigenstates, $\mathsf{B}^0_{(\mathsf{s})}$ and $\overline{\mathsf{B}}^0_{(\mathsf{s})}$ which do not correspond to the mass eigenstates, $\mathsf{B}^0_{(\mathsf{s}),\mathsf{L}}$ and $\mathsf{B}^0_{(\mathsf{s}),\mathsf{H}}$ in which they evolve in time. To study the interplay of the decay with this phenomenon, analyses are usually performed in bins of decay-time or imply flavour tagging techniques to identify the initial flavour, B or $\overline{\mathsf{B}}$, of the meson [38]. These types of analyses require signal samples much larger than what can be currently produced for $\mathsf{B}^0_\mathsf{s} \to \ell^+\ell^-$ at the LHC. Hence, only the time integrated and flavour-summed $\mathsf{B}^0_\mathsf{s} \to \ell^+\ell^-$ branching fraction is accessible at the time being.

However, in the case of the B^0_s meson this simple observable remains sensitive to the mixing and decay interplay [39, 40, 41]. For this meson, the two mass eigenstates $\mathsf{B}^0_{\mathsf{s},\mathsf{L}}$ and $\mathsf{B}^0_{\mathsf{s},\mathsf{H}}$ have different widths ($\Gamma_\mathsf{L} - \Gamma_\mathsf{H} = 0.0808^{+0.0077}_{-0.0078}$ ps [42]). Hence the time dependent decay rate is the sum of two decreasing exponential contributions:

$$\Gamma(\mathsf{B}^0_\mathsf{s}(t) \to \ell^+\ell^-) \;=\; \mathcal{N}\left[e^{-\Gamma_\mathsf{L} t}\langle \ell\ell|\mathsf{B}^0_{\mathsf{s},\mathsf{L}}\rangle + e^{-\Gamma_\mathsf{H} t}\langle \ell\ell|\mathsf{B}^0_{\mathsf{s},\mathsf{H}}\rangle\right], \qquad (1.12)$$

where $\mathcal{N}$ is a time independent factor. From Equation 1.12 the time integrated branching fraction can be written as:

$$\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \ell^+\ell^-) \;=\; \frac{1 + y_s \mathcal{A}_{\Delta\Gamma}}{1 - y_s^2} \times \mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \ell^+\ell^-)_{t=0}, \qquad (1.13)$$

with

$$y_s \;=\; \frac{\Gamma_\mathsf{L} - \Gamma_\mathsf{H}}{\Gamma_\mathsf{L} + \Gamma_\mathsf{H}} \qquad (1.14)$$

and

$$\mathcal{A}_{\Delta\Gamma} \;=\; \frac{\Gamma_{\mathsf{B}^0_{\mathsf{s},\mathsf{H}} \to \ell^+\ell^-} - \Gamma_{\mathsf{B}^0_{\mathsf{s},\mathsf{L}} \to \ell^+\ell^-}}{\Gamma_{\mathsf{B}^0_{\mathsf{s},\mathsf{H}} \to \ell^+\ell^-} + \Gamma_{\mathsf{B}^0_{\mathsf{s},\mathsf{L}} \to \ell^+\ell^-}}. \qquad (1.15)$$

Hence, via $\mathcal{A}_{\Delta\Gamma}$, the branching fraction would be affected in a non trivial way if new physics entered differently the $\mathsf{B}^0_{\mathsf{s},\mathsf{H}} \to \ell^+\ell^-$ and $\mathsf{B}^0_{\mathsf{s},\mathsf{L}} \to \ell^+\ell^-$ amplitudes.

Understanding the new physics responsible from an experimental discrepancy from the SM time integrated branching fraction prediction would require

a measurement of $\mathcal{A}_{\Delta\Gamma}$. The reader can find a description of a method to measure this observable based on the decay effective lifetime,

$$\tau_f \;=\; \frac{\int < \Gamma(\mathsf{B}(t) \to f) > t\mathrm{d}t}{\int < \Gamma(\mathsf{B}(t) \to f) > \mathrm{d}t}, \tag{1.16}$$

in Appendix A of [39]. Unfortunately, to be relevant, such a measurement needs a large signal sample. The $\mathsf{B}^0_\mathsf{s} \to \mathsf{K}^+\mathsf{K}^-$ effective lifetime measurement performed by LHCb [43][10] allows to derive prospects for $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$. With the full LHCb data set ($50\,\mathrm{fb}^{-1}$ corresponding to ~ 1000 $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ events if the branching fraction is SM-like) $\mathcal{A}_{\Delta\Gamma}$ for $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ would be measured with a precision of 75%.

For $\mathsf{B}^0 \to \ell^+\ell^-$ the width difference between the two mass eigenstates is negligible, hence $y_d \simeq 0$ and the branching fraction is no longer sensitive to the mixing and decay interplay. For more details on B mixing we report the reader to [37, 36] and Chapter 1 of [44].

As experiments are progressing tremendously in the search of $\mathsf{B}^0_{(\mathsf{s})} \to \mu^+\mu^-$ and their results are pointing close to the SM branching fraction values, precise predictions are needed, and the predicted quantities must correspond exactly to what is experimentally measured. For this second point, subtle effects due to the time dependency of the decay rate must be considered.

1.3.2 Branching Fraction Expression

The $\mathsf{B}^0_{(\mathsf{s})} \to \ell^+\ell^-$ branching fraction can be computed in a model independent way using the EFT with the following effective Hamiltonian [41]:

$$\mathcal{H}_{\mathsf{eff}} \;=\; -\frac{G_F\alpha}{\sqrt{2}\pi}\left\{ V_{\mathrm{tb}}V^*_{\mathrm{tq}} \sum_{i\in[10,S,P]} (C_i\mathcal{O}_i + C'_i\mathcal{O}'_i) + \mathrm{h.c.} \right\}, \tag{1.17}$$

where *h.c.* stands for hermitian conjugate, α is the QED fine structure constant, $C_i^{(\prime)}$ are the Wilson coefficients, and $\mathcal{O}_i$ the following local operators:

$$\begin{aligned} \mathcal{O}_{10} &= (\bar{q}\gamma^\mu P_L b)(\bar{\ell}\gamma_\mu\gamma_5\ell), \\ \mathcal{O}_S &= m_b(\bar{q}P_R b)(\bar{\ell}\ell), \\ \mathcal{O}_P &= m_b(\bar{q}P_R b)(\bar{\ell}\gamma_5\ell), \end{aligned} \tag{1.18}$$

[10] With 522 signal events and a signal to background ratio of around 10, the effective lifetime was measured with a precision of 7%.

with $P_{L,R}$ the right and left handed projection operators $(1 \mp \gamma_5)/2$. The primed operators $\mathcal{O}'_i$ correspond to the right handed current[11] and are obtained by exchanging P_R and P_L.

From this Hamiltonian, the time integrated untagged helicity-summed branching fraction is given by:

$$\begin{aligned}\mathcal{B}(\mathsf{B}^0_{(s)}(t) \to \ell^+\ell^-) &= \frac{\tau_{\mathsf{B}} G_F^4 M_{\mathrm{W}}^2 \sin^4\theta_W}{8\pi^5} \left|C_{10}^{\mathrm{SM}} V_{\mathrm{tb}} V_{\mathrm{tq}}^*\right|^2 F_B^2 m_B m_\ell^2 \\ &\times \sqrt{1 - \frac{4m_\ell^2}{m_{\mathsf{B}}^2}} \times (|P|^2 + |S|^2) \\ &\times \frac{1 + y_{s,d}\mathcal{A}_{\Delta\Gamma}}{1 - y_s^2}, \end{aligned} \tag{1.19}$$

where:

- τ_{B} is the $\mathsf{B}^0_{(s)}$ lifetime,
- G_F the Fermi constant,
- M_{W} the $\mathsf{W}^\pm$ boson mass,
- θ_W the Weinberg angle,
- F_B the $\mathsf{B}^0_{(s)}$ decay constants (discussed below),
- m_{B} the $\mathsf{B}^0_{(s)}$ mass,
- m_ℓ the lepton mass,
- S and P the Wilson coefficient combinations defined as:

$$\begin{aligned} P &= \frac{C_{10} - C'_{10}}{C_{10}^{\mathrm{SM}}} + \frac{m_{\mathsf{B}}^2}{2m_\ell} \frac{m_b}{m_b + m_q} \frac{C_P - C'_P}{C_{10}^{\mathrm{SM}}} = |P| e^{i\varphi_P}, \\ S &= \sqrt{1 - \frac{4m_\ell^2}{m_{\mathsf{B}}^2}} \frac{m_{\mathsf{B}}^2}{2m_\ell} \frac{m_b}{m_b + m_q} \frac{C_S - C'_S}{C_{10}^{\mathrm{SM}}} = |S| e^{i\varphi_S}, \end{aligned} \tag{1.20}$$

  to factorise in the branching fraction expression the SM contribution which only enters in C_{10} (see Section 1.4) and is noted C_{10}^{SM}. Hence, in the SM $P = 1$ and $S = 0$. Note that with this definition, $C_{10}^{(\prime)}$ are dimensionless, while $C_S^{(\prime)}$ and $C_P^{(\prime)}$ have the dimension of an inverse mass.

[11] The right handed current contribution is usually negligible in the SM.

The decay constants F_B account for the non-perturbative part of the matrix elements and are computed with lattice QCD. In the past they used to be the dominant source of uncertainty of the branching fraction and another method avoiding to use it was developed [45]. This method is based on the meson mass difference Δm_{B}, which in the minimal flavour violation hypothesis (see Section 1.5.2) is proportional to F_B and B_{B}, some renormalisation group invariant parameter. This parameter used to be better known and controlled than F_B. Thus, by using the measured value of Δm_{B}, the branching fraction could be predicted with a better precision than in the method using F_B. However the F_B based method is now preferred as the B_{B} based method is valid only in minimal flavour violation hypothesis and recent progresses in lattice computation allow to predict F_B with a precision of 1.3% [46]. Waiting for confirmation of these impressive progresses the conservative approach is to use a simple average of the different F_B predictions (summarised in [47]) for the central value and set the uncertainty to a reasonable value [48] which gives:

$$F_{\mathsf{B}^0_\mathsf{s}} = (227 \pm 8)\,\text{MeV} \tag{1.21}$$
$$F_{\mathsf{B}^0} = (190 \pm 7)\,\text{MeV}. \tag{1.22}$$

Finally, for $\mathsf{B}^0_\mathsf{s} \to \ell^+\ell^-$, $\mathcal{A}_{\Delta\Gamma}$ is given by:

$$\mathcal{A}_{\Delta\Gamma} = \frac{|P|^2\cos(2\varphi_P - \phi_s^{NP}) - |S|^2\cos(2\varphi_S - \phi_s^{NP})}{|P|^2 + |S|^2}, \tag{1.23}$$

where ϕ_s^{NP} is the CP-violating phase contribution due to physics beyond the SM in the B^0_s mixing.

1.4 Standard Model Predictions

Based on the general expression of the branching fraction and $\mathcal{A}_{\Delta\Gamma}$ developed in the previous section, SM predictions can be derived. Section 1.4.1 describes the SM contributions and explains the essential points to obtain precise and reliable predictions. Finally, Section 1.4.2 and Section 1.4.3 give the SM predictions for $\mathsf{B}^0_{(\mathsf{s})} \to \mu^+\mu^-$ and $\mathsf{B}^0_{(\mathsf{s})} \to \tau^+\tau^-$.

1.4.1 Standard Model Contribution Evaluation

The two quantities to be computed to obtain the SM branching fraction predictions are C_{SM}^{10} and $\mathcal{A}_{\Delta\Gamma}$ as shown in Equation 1.19. $\mathcal{A}_{\Delta\Gamma}$ can be evaluated from Equation 1.23 with $S = 0$, $P = 1$, and $\phi_s^{\text{NP}} = 0$ which gives $\mathcal{A}_{\Delta\Gamma} = 1$.

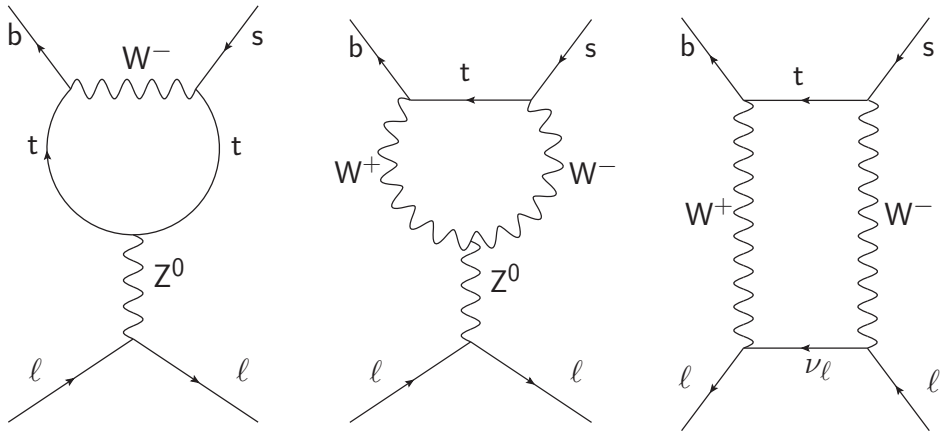


Figure 1.6: $\mathsf{B}^0_{(s)} \to \ell^+\ell^-$ penguins (a,b) and box (c) dominant SM Feynman diagrams.

$C^{10}_{\rm SM}$ is obtained by evaluating the contributions coming from the Z^0 penguin and $\mathsf{W}^\pm$ box diagrams shown in Figure 1.6.

After the recent discovery of a SM-Higgs like boson by the ATLAS and CMS collaborations lets mention that some contributions from the SM Higgs boson also enter the C_{10} Wilson coefficient [32]. Since the Higgs coupling are proportional to the fermion mass, the only contributions which are substantial are those for which the Higgs boson couples at both ends of its propagator to the top. The main Feynman diagrams to be considered are shown in Figure 1.7 and their contribution is below the percent level. Hence $\mathsf{B}^0_s \to \ell^+\ell^-$ processes are not sensitive to the Higgs SM sector.

The computation of C_{10} is performed in the framework of the EFT which implies an arbitrary choice of energy scale (μ) and renormalisation scheme. As explained in Section 1.2, it is crucial to check that the final amplitude is independent of these choices.

The energy scale dependency is strongly reduced by including next to leading order (NLO) QCD corrections (i.e. diagrams like in Figure 1.7 replacing the Higgs by a gluon) [48, 49, 50] as shown in Figure 1.8.

On the other hand the choice of renormalisation scheme is still an open issue. The ultimate solution of this problem requires a full NLO electroweak calculation but nowadays only a two loop electroweak computation is available [51]. However such a full NLO electroweak calculation was performed for $\mathsf{K}^+ \to \pi^+\nu\bar{\nu}$ [52] and allowed to choose the correct scheme. Fortunately $\mathsf{K}^+ \to \pi^+\nu\bar{\nu}$ involves the same penguins diagrams as $\mathsf{B}^0_{(s)} \to \ell^+\ell^-$. Hence it is

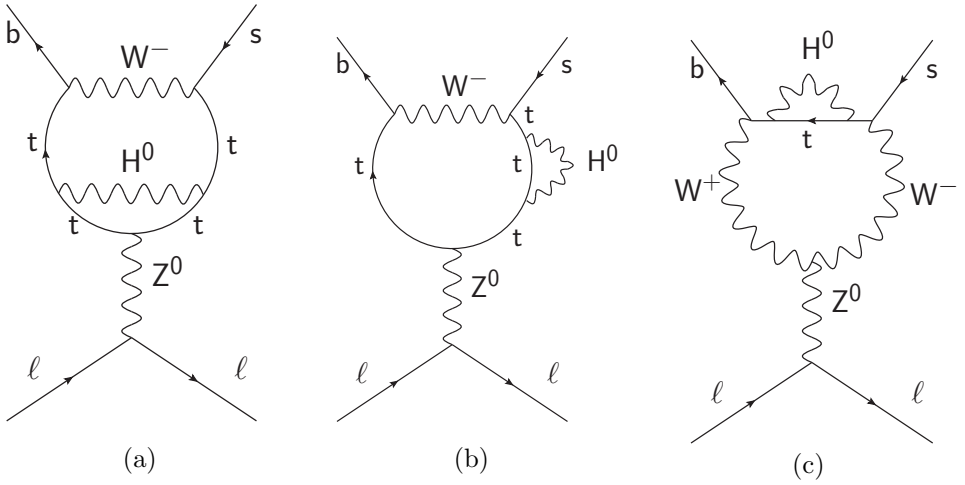


Figure 1.7: Main $\mathsf{B}^0_{(s)} \to \ell^+\ell^-$ Feynman diagrams with a SM Higgs boson.

safe to assume that the same renormalisation scheme can be used [48]. After having tested the independence on the energy scale and renormalisation scheme, $C^{10}_{\rm SM}$ can be computed with a precision below 1% [48].

1.4.2 $\mathsf{B}^0_{(s)} \to \mu^+\mu^-$ SM Predictions

Merging all information from the previous section, the initial flavour averaged branching fractions in the SM are:

$$\mathcal{B}(\mathsf{B}^0_s(t) \to \mu^+\mu^-)_{t=0} = (3.35 \pm 0.28) \times 10^{-9} \quad (1.24)$$

$$\mathcal{B}(\mathsf{B}^0 \to \mu^+\mu^-) = (1.07 \pm 0.10) \times 10^{-10}. \quad (1.25)$$

This prediction is obtained by updating the results from [48] with the most up to date $\tau_{\mathsf{B}^0_s}$ experimental average ($\tau_{\mathsf{B}^0_s} = (1.5185 \pm 0.0068)$ ps [42]). The time integrated $\mathsf{B}^0_s \to \mu^+\mu^-$ branching fraction is obtained by dividing Equation 1.24 by $(1 - y_s)$ since $\mathcal{A}_{\Delta\Gamma} = 1$ in the SM. The latest experimental average for this parameter is $y_s = 0.0613 \pm 0.0059$ [42]. This value is correlated with $\tau_{\mathsf{B}^0_s}$ which enters the computation of Equation 1.24. Accounting for this correlation the time integrated $\mathsf{B}^0_s \to \mu^+\mu^-$ branching fraction is:

$$\mathcal{B}(\mathsf{B}^0_s \to \mu^+\mu^-) = 3.57 \pm 0.30 \times 10^{-9}. \quad (1.26)$$

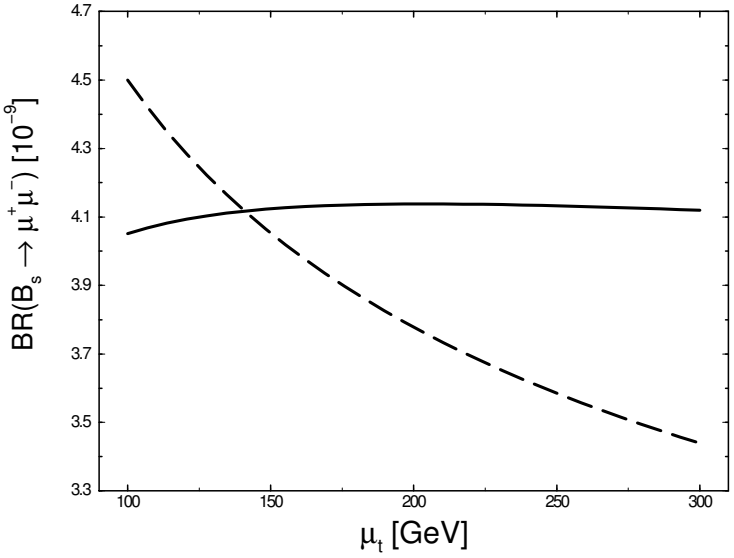


Figure 1.8: Energy scale dependency for $\mathcal{B}(\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-)$. The solid and dashed curves denote the predictions with and without QCD corrections, respectively [50].

1.4.3 $\mathsf{B}^0_{(\mathsf{s})} \to \tau^+\tau^-$ SM Predictions

The branching fraction predictions for $\mathsf{B}^0_{(\mathsf{s})} \to \tau^+\tau^-$ in the SM, are simply derived from Equations 1.25 and 1.26 with the following relation:

$$\frac{\mathcal{B}(\mathsf{B}^0_{(\mathsf{s})} \to \tau^+\tau^-)}{\mathcal{B}(\mathsf{B}^0_{(\mathsf{s})} \to \mu^+\mu^-)} = \frac{m_\tau^2}{m_\mu^2} \times \sqrt{\frac{m_\mathsf{B}^2 - 4m_\tau^2}{m_\mathsf{B}^2 - 4m_\mu^2}}. \tag{1.27}$$

As a rule of thumb we can remember that the ratio of the two branching fractions is around 210. More precisely the time integrated branching fractions are:

$$\mathcal{B}(\mathsf{B}_\mathsf{s}^0 \to \tau^+\tau^-) = (7.57 \pm 0.63) \times 10^{-7} \tag{1.28}$$
$$\mathcal{B}(\mathsf{B}^0 \to \tau^+\tau^-) = (2.24 \pm 0.10) \times 10^{-8}. \tag{1.29}$$

1.5 Phenomenology of $\mathsf{B}_\mathsf{s}^0 \to \ell^+\ell^-$ Beyond the Standard Model

In the previous sections we have introduced several observables. The most important are the time integrated and flavour averaged branching fractions

of $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$, $\mathsf{B}^0 \to \mu^+\mu^-$, $\mathsf{B}^0_\mathsf{s} \to \tau^+\tau^-$ and $\mathsf{B}^0 \to \tau^+\tau^-$. In this section we will try to explain how these observables allow to test theories beyond the SM and shed light on their complementarity. After a general discussion, in Section 1.5.1, on the sensitivity of these modes to BSM physics from the view point of the Wilson coefficients, we will describe, in Section 1.5.2, the phenomenology of few selected models in more details.

In Section 1.3 we also introduced another observable $\mathcal{A}_{\Delta\Gamma}$, as it enters the $\mathsf{B}^0_\mathsf{s} \to \ell^+\ell^-$ time integrated branching fraction expression. In order to measure this observable, a fit to the effective lifetime is required [40]. Unfortunately the statistics needed to perform a time dependent analysis will be available only at the end of the LHCb upgrade. However, $\mathcal{A}_{\Delta\Gamma}$ is essential to parametrise the signal lifetime distribution of the simulated data which are used in the current analysis. As a consequence, the experimental results depend on $\mathcal{A}_{\Delta\Gamma}$ (this point will be discussed in details in Section 4). In order to extract the most stringent constraints from the data it is important to compare the predictions and measurements both as a function of $\mathcal{A}_{\Delta\Gamma}$. Hence in the description of the various phenomenologies beyond the SM, we will try as much as possible to give the branching fraction prediction as a function of $\mathcal{A}_{\Delta\Gamma}$.

1.5.1 Model Independent Discussion

The sensitivity of the branching fraction to BSM physics can be viewed from its dependency on the Wilson coefficients[12]:

$$\begin{aligned}\mathcal{B}(\mathsf{B}^0_{(\mathsf{s})}(\mathsf{t}) \to \ell^+\ell^-)_{t=0} \;&\propto\; \left(1 - \frac{4m_\ell^2}{m_\mathsf{B}^2}\right)|C_S - C'_S|^2 \\ &+\; \left|(C_P - C'_P) + 2\frac{m_\ell}{m_\mathsf{B}^2}(C_{10} - C'_{10})\right|^2 . \quad (1.30)\end{aligned}$$

As mentioned in Section 1.4, the SM contribution only enters C_{10}. Equation 1.30 shows that all the contributions entering $C_{10}^{(\prime)}$ are multiplied by the helicity suppression factor proportional to the lepton mass m_ℓ. On the contrary the contribution entering $C_{S,P}^{(\prime)}$ do not suffer from this suppression factor. Hence the branching fractions of $\mathsf{B}^0_{(\mathsf{s})} \to \ell^+\ell^-$ are very sensitive to the models which could affect the scalar and pseudo-scalar Wilson coefficients like theories with extended Higgs sectors.

[12]Only the sensitivity of the initial CP averaged branching fraction is discussed here. The sensitivity of $\mathcal{A}_{\Delta\Gamma}$ will be discuss later for specific models.

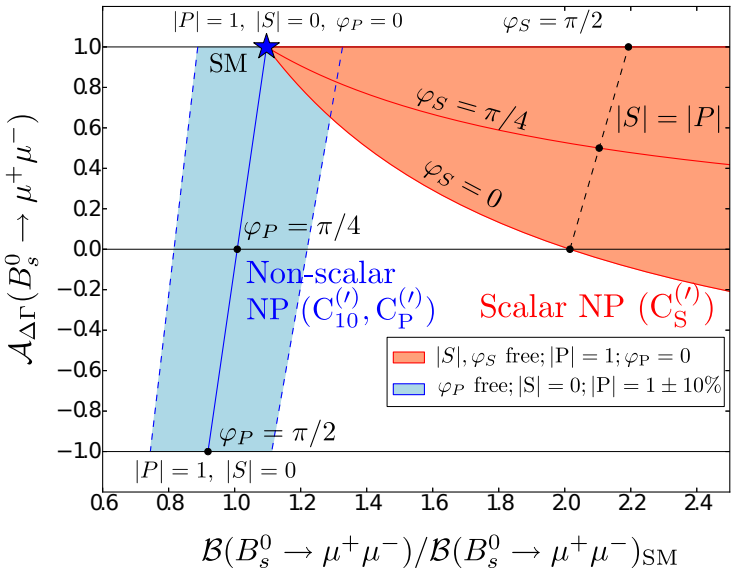


Figure 1.9: Allowed region in the plane $\mathcal{B}(\mathsf{B_s^0}\to\mu^+\mu^-)/\mathcal{B}(\mathsf{B_s^0}\to\mu^+\mu^-)_{\rm SM}\times\mathcal{A}_{\Delta\Gamma}$ if new physics contributions are or not scalar [40]. S, P, $\varphi_{S,P}$ are defined in Equation 1.20.

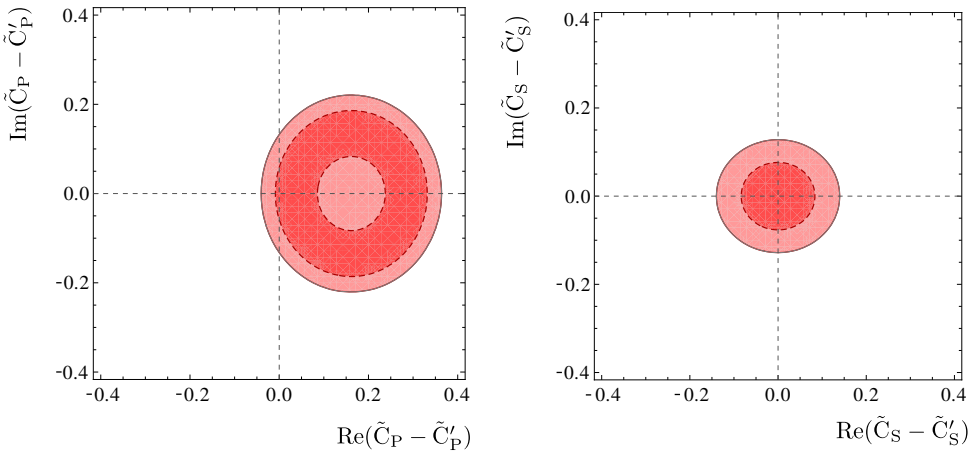
Scalar and Pseudo-scalar Contribution Phenomenologies

New scalar contributions enter as $|C_S - C_S'|^2$ in the branching fraction expression, hence they necessarily enhance it. On the contrary pseudo-scalar contributions could enhance but also suppress the branching fraction as they can interfere destructively when they are summed with $C_{10}^{\rm SM}$ (see Equation 1.30). Figure 1.9 illustrates these two different phenomenologies and shows as well that the scalar and non-scalar contributions have very different $\mathcal{A}_{\Delta\Gamma}$ dependency[13].

Lacking hints of BSM physics, constraints can still be set on the Wilson coefficients. Figure 1.10 shows the constraints which can be set from a combined branching fraction estimate based on the CDF and LHC 2012 $\mathsf{B_s^0}\to\mu^+\mu^-$ results [53, 54, 55, 56]. Note that the scalar and pseudo-scalar Wilson coefficient definitions employed in this Figure and in Equation 1.19 are different. They are related by:

$$\tilde{C}_{S,P}^{(\prime)} \;=\; C_{S,P}^{(\prime)}/m_{\mathsf{B_s^0}}. \tag{1.31}$$

[13] In Figure 1.9, $C_{10}^{(\prime)}$ contributions are also allowed.

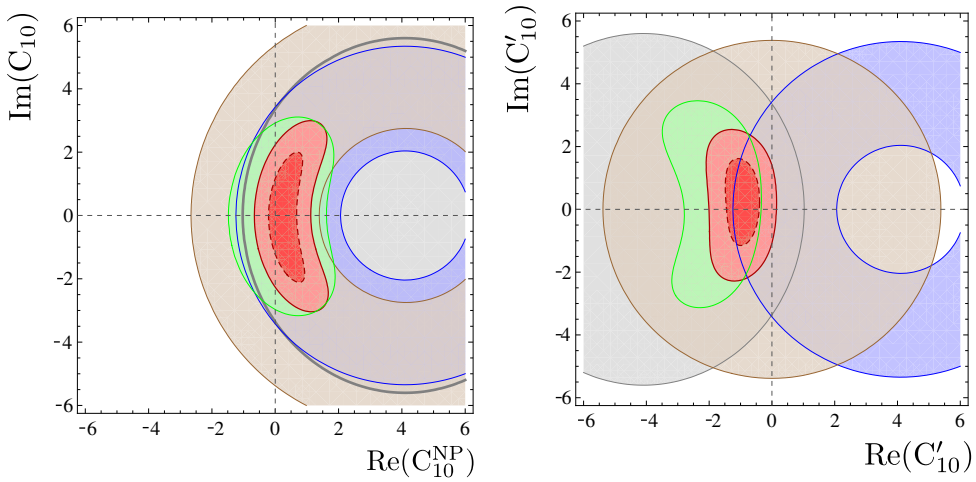


(a) (b)

Figure 1.10: Values in the plane $\mathrm{Re}(\tilde{C}_S - \tilde{C}'_S) \times \mathrm{Im}(\tilde{C}_S - \tilde{C}'_S)$ (a) and $\mathrm{Re}(\tilde{C}_P - \tilde{C}'_P) \times \mathrm{Im}(\tilde{C}_P - \tilde{C}'_P)$ (b) consistent at 1 σ with $\mathcal{B}(\mathsf{B}^0_s \to \mu^+\mu^-) = 2.4 \pm 1.6 \times 10^{-9}$ if $C_{10}^{(\prime)}$ is purely SM [26]. $\tilde{C}_{S,P}^{(\prime)}$ are defined in Equation 1.31.

$C_{10}^{(\prime)}$ Contribution Phenomenology

Some contributions can also enter $(C_{10} - C'_{10})$ in models with a fourth fermion generations, extra dimensions (e.g. Randall-Sundrum) or composite models (e.g. Litte Higgs Model with T-parity or Topcolour-assisted Technicolour). Despite the above mentioned helicity suppression, some useful constraints on these models can still be obtained, in particular using the interplay with $\mathsf{B}^0 \to \mathsf{K}^{*0}\mu^+\mu^-$ observables which are sensitive to the Wilson coefficient sum $(C_{10} + C'_{10})$. In the case of no substantial new physics scalar contribution, $C_{10}^{(\prime)\mathrm{NP}}$ could enhance or suppress the branching fraction by interfering constructively or destructively with C_{10}^{SM} (see Equation 1.30), as shown by the blue band in Figure 1.9. Again, lacking hints of BSM physics, constraints can be set on these coefficients. Figure 1.11 [26] shows the constraints brought by several channels as well as their combinations and demonstrates that, despite the helicity suppression affecting the $C_{10}^{(\prime)}$ contributions, upper limits on $\mathcal{B}(\mathsf{B}^0_s \to \mu^+\mu^-)$ bring some of the strongest constraints. The stringent combined constraints also illustrate the comment made in Section 1.2 on the importance of the observable interplay which can be exploited thanks to the EFT framework



(a) (b)

Figure 1.11: Values for the C_{10} (a) and C'_{10} (b) Wilson coefficients consistent at 1 σ with the experimental results from $B^0 \to X_s \ell^+ \ell^-$ (brown), $B^0 \to K \mu^+ \mu^-$ (blue), $B^0 \to K^{*0} \mu^+ \mu^-$ (green), $\mathcal{B}(B^0_s \to \mu^+ \mu^-) = 2.4 \pm 1.6 \times 10^{-9}$ (gray) and with their combination (red) at 1 and 2 σ [26].

1.5.2 Model Dependent Discussion

This section proposes a quick tour of some of the theories beyond the SM and of the phenomenologies they predict for $\mathcal{B}(B^0_s \to \ell^+ \ell^-)$. This review is far from being exhaustive and the models' phenomenology is not discussed in technical terms. The choice of the models presented was made to emphasise the complementarity between the different $B^0_{(s)} \to \ell^+ \ell^-$ modes and the importance of the branching fraction dependency on $\mathcal{A}_{\Delta\Gamma}$.

Minimal Flavour Violation Hypothesis

The Minimal Flavour Violation (MFV) is a hypothesis emerged in the 1980's (see references in [57]) postulating that the dynamics of flavour and CP violation in physics beyond the SM is governed by the known Yukawa structure. The hypothesis gained in credibility as experimental results strongly constrained the possibility for new physics to couple to quarks with a non-Yukawa structure. Recently a precise formulation of MFV was developed [58] based on the EFT approach. Most models described in the following respect this hypothesis.

One of the most stringent tests of MFV is the ratio of the $B^0_s \to \ell^+ \ell^-$ and $B^0 \to \ell^+ \ell^-$ branching fractions, which has a very clean expression under this

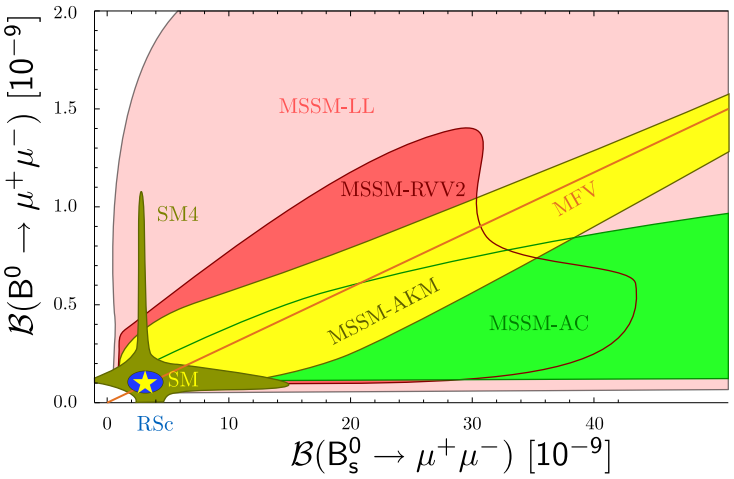


Figure 1.12: Prediction for the branching fractions of $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ in Minimal Flavour Violation models (MFV, see Section 1.5.2), the SM with a fourth generation (SM4, see Section 1.5.2), the Randall-Sundrum model (RSc) with custodial protection [60], and four Minimal SUSY favour models (MSSM), Agashe and Carone (AC) [61], Ross, Velasco-Sevilla and Vives (RVV2) [62], Antusch, King and Malinsky (AKM10) [63] and a model with left-handed currents only (LL11) [64]. The SM prediction is marked by a star [59].

hypothesis [57]:

$$\frac{\mathcal{B}(B^0 \to \ell^+\ell^-)}{\mathcal{B}(B_s^0 \to \ell^+\ell^-)} = \frac{\tau_{B^0}}{\tau_{B_s^0}} \frac{m_{B^0}}{m_{B_s^0}} \frac{F_{B^0}}{F_{B_s^0}} \left|\frac{V_{td}}{V_{ts}}\right|^2 . \tag{1.32}$$

Figure 1.12 [59] shows this relation between the two $B_{(s)}^0 \to \mu^+\mu^-$ branching fractions under the MFV hypothesis.

Two Higgs Doublet Model

The Higgs mechanism with one doublet is in fact the simplest model to describe electroweak symmetry breaking. However more sophisticated models can be introduced like the two Higgs doublet models (2HDMs) where the Higgs sector contains two doublets. From these doublets, five Higgs particles arise, two charged scalar Higgs, $H^\pm$, two neutral scalars, H^0 and h^0, and a neutral pseudo-scalar, A^0.

The vacuum expectation values of the two doublets, v_1 and v_2, allow to define an angle, β, such that $\tan\beta = v_2/v_1$. This angle is one of the free parameters of the theory.

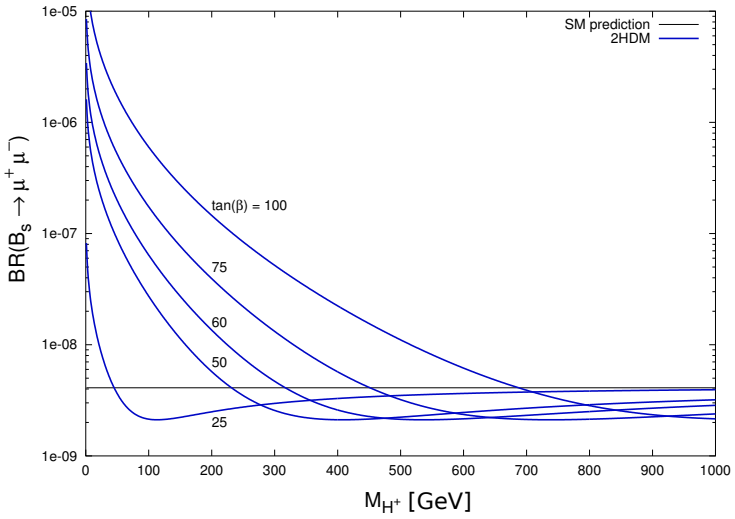


Figure 1.13: $\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-)$ in the 2HDM-II as a function of $M_{\mathsf{H}^\pm}$ for $\tan\beta =$ 100, 75, 60, 50, and 25 [65].

The 2HDMs are categorised in four types according to the coupling between the Higgs fields and the fermions. In the following we will focus only on the type 2 (2HDM-II), where the first doublet couples to down-type fermions and the second to up-type. This model is of particular interest as it predicts $\mathsf{B}^0_{(\mathsf{s})} \to \ell^+\ell^-$ branching fractions significantly enhanced if $\tan\beta$ is large. Note that the Minimal Super-symmetric extension of the SM, described in the following paragraph falls into this category.

In the 2HDM-II, the $\mathsf{B}^0_\mathsf{s} \to \ell^+\ell^-$ branching fractions get scalar and pseudo scalar contributions proportional to $\tan^4\beta$. Hence if $\tan\beta$ is large, the branching fraction is substantially enhanced. This enhancement is all the more significant as the scalar and pseudo-scalar contributions do not suffer from the helicity suppression affecting the SM. Typical branching fractions values for $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ in the 2HDM-II are shown in Figure 1.13.

$\mathcal{A}_{\Delta\Gamma}$ is also affected by these new contributions. As explained in Section 1.3, the time integrated branching fraction and $\mathcal{A}_{\Delta\Gamma}$ are correlated in a non-trivial way. In Figure 1.14, this correlation is shown for $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ in the decoupled 2HDM-II scenario [41]. In this scenario the mass of the heavy scalar and pseudo-scalar, H^0 and A^0, are much larger than that of the light scalar h^0 and therefore $C_S^{(\prime)} = \pm C_P^{(\prime)}$. As shown in Figure 1.14, this scenario is characterised by the lower bound $\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-)/\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-)_{\rm SM} > (1 - y_s)/2$.

Finally, even though the $\mathsf{B}^0_{(\mathsf{s})} \to \tau^+\tau^-$ modes are challenging experimen-

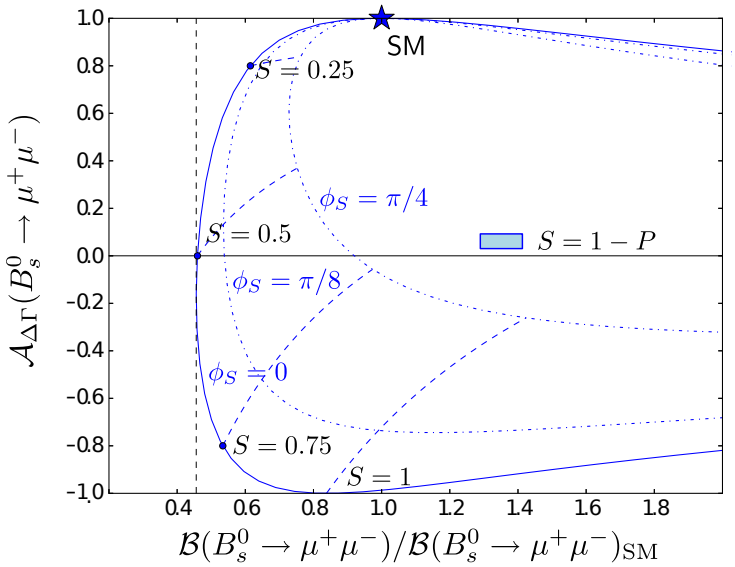


Figure 1.14: Correlations between the time integrated branching fraction, normalised to the SM prediction, and $\mathcal{A}_{\Delta\Gamma}$ observables. The solid and dashed blue lines give the prediction without and with violating phases [41]. S, P, $\varphi_{S,P}$ are defined in Equation 1.20.

tally, they are theoretically more interesting than $\mathsf{B}^0_{(s)} \to \mu^+\mu^-$ as they can probe features that are inaccessible to the light leptons modes. Indeed [66], all contributions entering with $m_{\mathsf{B}} \tan\beta m_\ell / m^2_{\mathsf{W}^\pm}$ as a prefactor can only be accessed with $\mathsf{B}^0_{(s)} \to \tau^+\tau^-$ as the large τ mass makes the prefactor non negligible.

Super-Symmetry

Super-symmetric models are extensions of the SM in which all fermions have boson superpartners and vice-versa. The resulting theories are therefore invariant under a new deep symmetry, called super-symmetry (SUSY), which transforms fermions into bosons. This hypothesis has the virtue to solve some of the SM issues. Indeed, in the computation of the Higgs masses, divergences due to radiative corrections cancel between super-partners and the Higgs mass is stabilised. In addition, in many scenarios, one of the super-partners is neutral, massive, and stable. These properties make it a possible dark matter candidate.

If SUSY was unbroken at our energies, super-partners would have the same masses. Since none of these new particles have been observed, SUSY

must have undertaken a spontaneous symmetry breaking which split the superpartners masses. Many solutions are proposed for this symmetry breaking and therefore many SUSY models have been developed.

The minimal super-symmetric extension of the SM (MSSM) is one of the most studied scenario. At tree level, it presents a Higgs sector with two doublets and couplings similar to a 2HDM-II model. Hence, as in the 2HDM-II, large enhancements can be predicted due to scalar and pseudoscalar contributions together with a non trivial correlation between $\mathcal{B}(\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-)$ and $\mathcal{A}_{\Delta\Gamma}$. However, besides the new Higgs fields, new superpartners are also added which can bring new contributions to $C_{10}^{(\prime)}$. As shown in Section 1.5.1, these contributions can enhance, but also reduce, the branching fractions compared to the SM predictions. Some predictions obtained for constrained versions of the MSSM are shown in Figure 1.12.

Fourth Generation

The question of the number of fermion families has been explored by the LEP measurements at the Z^0-pole [67] and pointed clearly to three families. However this interpretation relied on the assumption of massless or light neutrinos and is no longer valid after the observation of neutrino mixing [5]. Hence the possibility of a fourth fermion family is still open, provided that the mass of the new neutrino be larger than $m_{\mathsf{Z}^0}/2$.

Nevertheless, the recently discovered Higgs boson almost excludes the fourth generation hypothesis if this boson is the SM one [68, 69]. In this context, the study of the $\mathsf{B}^0 \to \mu^+\mu^-$ decay plays a crucial role [70]. Indeed, as shown in Figure 1.12, the fourth generation hypothesis has the specific feature of being able to enhance the branching fraction of only one of the $\mathsf{B}_\mathsf{s}^0 \to \ell^+\ell^-$ or $\mathsf{B}^0 \to \ell^+\ell^-$ modes. The branching fraction of $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ being already strongly constrained, an enhancement of the branching fraction of $\mathsf{B}^0 \to \mu^+\mu^-$ could point towards a fourth generation and indicate that the newly observed Higgs boson may not be the SM one [70].

The $\mathcal{A}_{\Delta\Gamma}$ observable, also in this case, is correlated with the branching fraction. Figure 1.15 shows this correlation for a generic class of models with $S = 0$ (S defined in Equation 1.20) in which falls the fourth fermion family hypothesis.

Leptoquarks

The apparent symmetry between quarks and leptons suggests that there could be a connection between them. The leptoquarks would be the bosons related to this connection and therefore carry both lepton and baryon num-

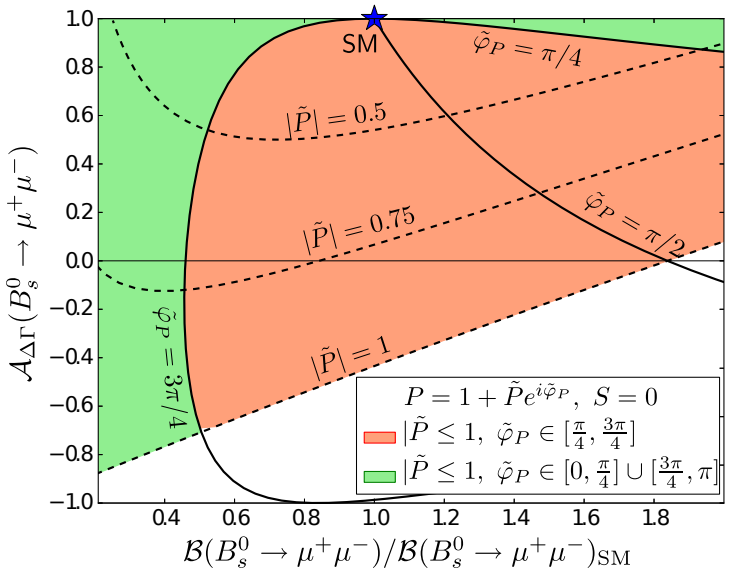


Figure 1.15: Allowed region in the plane $\mathcal{B}(\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-)/\mathcal{B}(\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-)_{\rm SM} \times \mathcal{A}_{\Delta\Gamma}$. S, P, $\varphi_{S,P}$ are defined in Equation 1.20 [41].

bers. The leptoquark hypothesis is realised in several models like Grand Unification Theories[14] (GUTs) and composite models[15]. Since they carry both lepton and baryon numbers, leptoquarks allow to directly couple quarks to leptons. Hence the $\mathsf{B}_{(\mathsf{s})}^0 \to \ell^+\ell^-$ processes could be mediated at tree level by leptoquarks, as shown in Figure 1.16.

However the coupling between the leptoquarks and the leptons depends on the lepton generation. Hence the study of the decay to di-muon and di-tau final states are complementarity.

Z' models

In Section 1.1.3 we showed that the electromagnetic and weak forces could be unified under the same interaction. Many theories try to extend the SM by operating a similar unification between the electroweak and the strong forces. To achieve this unification, the SM gauge group needs to be replaced and the new gauge group would have a rank larger than in SM hence, new gauge bosons are expected. From a phenomenological point of view, the introduction of a new neutral colorless spin-1 gauge boson Z', could strongly

[14] In GUTs the weak, strong, and gravitational interaction are unified.

[15] In composite models the SM particle are supposed to be composed. The substructures corresponding to leptoquarks are called preons.

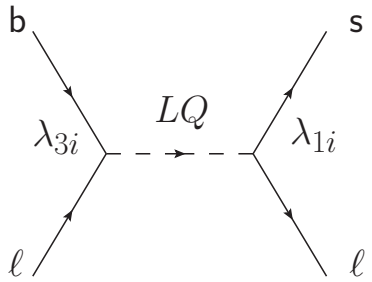


Figure 1.16: Feynman diagram for $\mathsf{B}^0_\mathsf{s} \to \ell^+\ell^-$ mediated by a leptoquark. λ_{ij} are the coupling of the leptoquark between the i and j families.

affect $\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \ell^+\ell^-)$ as, if the Z' couplings are flavour violating, these processes could be mediated at tree levels. A general overview of these type of models is provided in [71, 72] and the phenomenology they imply for flavour changing neutral currents is extensively described in [73].

As far as $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ is concerned, both the branching fraction and $\mathcal{A}_{\Delta\Gamma}$ would be affected in a non trivial way. Figure 1.17 shows the correlated prediction for these two observables if only left-handed current were allowed [41].

In the case of $\mathsf{B}^0_{(\mathsf{s})} \to \tau^+\tau^-$ the branching fraction could be enhanced up to the percent level. This observable is all the more appealing [74, 75] as the new physics able to enhance $\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \tau^+\tau^-)$ could also contribute to the absorptive part of the B^0_s mixing, which could explain the large like-sign di-muon asymmetry observed by D0 [76, 77, 78][16].

[16] LHCb recent analyses [79] measured an asymmetry which is compatible with the SM but does not exclude the D0 results.

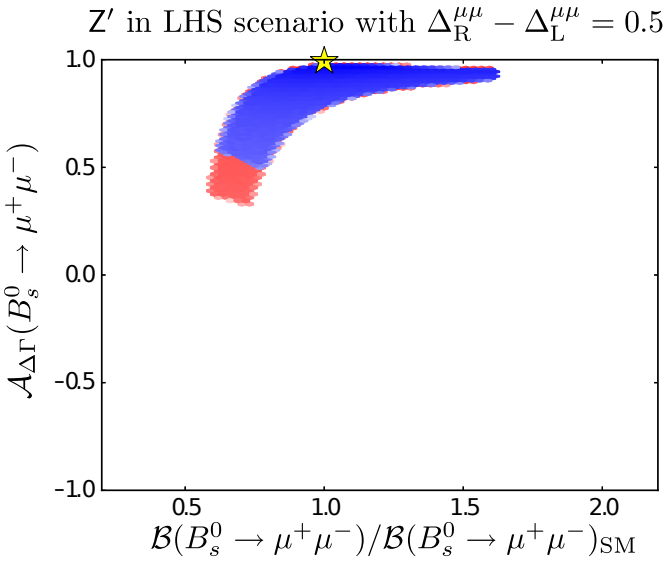


Figure 1.17: Allowed region in the plane $\mathcal{B}(\mathrm{B_s^0} \to \mu^+\mu^-)/\mathcal{B}(\mathrm{B_s^0} \to \mu^+\mu^-)_{\mathrm{SM}} \times \mathcal{A}_{\Delta\Gamma}$ assuming only left-handed currents, $m_{\mathrm{Z'}} = 1\,\mathrm{TeV}$, and $\mathrm{Z'}$ couplings between muons related as $\Delta_R^{\mu\mu} - \Delta_L^{\mu\mu} = 0.5$. The two colours represent two oases of allowed values for the couplings. The star is the SM prediction [41].

1.6 Conclusion

This chapter introduced the theory currently used to describe high energy physics and called Standard Model (SM). After exposing its main fundamental concepts, we showed that, for several experimental and theoretical reasons, the SM may only be an effective low energy version of a more general theory. Probing this general theory means therefore probing high energy processes. This type of processes can be accessed through heavy flavour meson decays. Nevertheless, in these decays the phenomenologies of high and low energy processes are entangled but can be separated by means of an effective field theory.

$\mathsf{B}^0_{(s)} \to \ell^+\ell^-$ are among to best heavy flavour meson decays to probe new physics. Indeed, the SM contributions to these decay amplitudes are small as they cannot be mediated at tree level in the SM. These contributions are all the more limited as these processes are helicity suppressed. Thus, new physics could easily bring contributions as large as the SM and modify the branching fraction up to an observable level.

In addition to the branching fraction another clean observable, $\mathcal{A}_{\Delta\Gamma}$ can be measured in $\mathsf{B}^0_s \to \ell^+\ell^-$ modes as the decay width difference between the B^0_s mass eigenstates is not zero. Even though a direct measurement of $\mathcal{A}_{\Delta\Gamma}$ will only be feasible at the LHCb upgrade, the correlations between the theoretical predictions for this observable and the branching fraction must already be considered. Indeed, we will show that the experimental branching fraction also depends on $\mathcal{A}_{\Delta\Gamma}$ and that some sensitivity to new physics can be gained by comparing experimental results and theoretical predictions for the branching fraction as a function of $\mathcal{A}_{\Delta\Gamma}$.

In the SM, the predictions for the $\mathsf{B}^0_{(s)} \to \ell^+\ell^-$ branching fractions are:

$$\mathcal{B}(\mathsf{B}^0_s \to \mu^+\mu^-) = (3.57 \pm 0.30) \times 10^{-9}, \quad (1.33)$$

$$\mathcal{B}(\mathsf{B}^0 \to \mu^+\mu^-) = (1.07 \pm 0.10) \times 10^{-10}, \quad (1.34)$$

$$\mathcal{B}(\mathsf{B}^0_s \to \tau^+\tau^-) = (7.57 \pm 0.63) \times 10^{-7}, \quad (1.35)$$

$$\mathcal{B}(\mathsf{B}^0 \to \tau^+\tau^-) = (2.24 \pm 0.10) \times 10^{-8}. \quad (1.36)$$

Beyond the standard model, the phenomenology can be described in two ways. First the effective field theory allows to hold a model independent discussion and classify new physics contributions. Such a discussion shows that the $\mathsf{B}^0_{(s)} \to \ell^+\ell^-$ modes are remarkably sensitive to scalar and pseudoscalar contributions as these contributions escape the helicity suppression affecting the SM. While scalar contributions necessarily implies branching fraction enhancement, other type of new physics contributions could interfere constructively or destructively with the SM contribution and reduce the

branching fraction. Then, model dependent discussions can be developed as they shed light on the complementarity between the various modes. For instance, the ratio of the $B_s^0 \to \ell^+\ell^-$ and $B^0 \to \ell^+\ell^-$ branching fractions appears to be a very clean observable and a crucial test for the minimal flavour violation hypothesis. In addition, the $B_{(s)}^0 \to \tau^+\tau^-$ and the $B_{(s)}^0 \to \mu^+\mu^-$ decays are complementary as new physics, in particular lepto-quarks, could feature lepton family dependent couplings.

The solid theoretical grounds for which $B_{(s)}^0 \to \ell^+\ell^-$ decays are commonly considered portals towards new physics, have motivated experimental searches since 1984 [80]. Since then, the experimental sensitivity to these decays has constantly increased and allows today, with the LHCb experiment, to start measuring the very feeble SM $B_s^0 \to \mu^+\mu^-$ signal. Such a breakthrough is possible with the LHCb detector thanks to a design partly dedicated to the search for $B_s^0 \to \mu^+\mu^-$. This detector is described in the next chapter.

Bibliography

[1] D. Griffiths. *Introduction to elementary particles* (2008).

[2] B. Hatfield. *Quantum field theory of point particles and strings*. Frontiers in Physics, Redwood City, CA, **1** (1992).

[3] J. S. Hagelin. *Is Consciousness the Unified Field?: a Field Theorist's Perspective* (1987).

[4] M. Strassler. *Virtual Particles: What are they?* Accessed May 25, 2013.

[5] Super-Kamiokande Collaboration. *Measurements of the solar neutrino flux from Super-Kamiokande's first 300 days*. Physical Review Letter, **81** (1998), 1158–1162.

[6] J. Iliopoulos. *An Introduction to Gauge Theories*. CERN-76-11 (1976).

[7] F. Englert et al. *Broken Symmetry and the Mass of Gauge Vector Mesons*. Physical Review Letter, **13** (1964) 9, 321–323.

[8] P. Higgs. *Broken symmetries, massless particles and gauge fields*. Physics Letter, **12** (1964) 2, 132–133.

[9] G. S. Guralnik, et al. *Global Conservation Laws and Massless Particles*. Physical Review Letter, **13** (1964) 20, 585–587.

[10] ATLAS Collaboration. *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*. Physics Letter, **B716** (2012), 1–29.

[11] CMS Collaboration. *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*. Physics Letter, **B716** (2012), 30–61.

[12] ATLAS and CMS Collaborations. *Higgs related talks*. Rencontres de Moriond, Electroweak Interactions and Unified Theories.

[13] J. Ellis et al. *Updated Global Analysis of Higgs Couplings*. Journal of High Energy Physics, **1306** (2013), 103.

[14] N. Cabibbo. *Unitary Symmetry and Leptonic Decays*. Physical Review Letter, **10** (1963) 12, 531–533.

[15] M. Kobayashi et al. *CP Violation in the Renormalizable Theory of Weak Interaction*. Progress of Theoretical Physics, **49** (1973), 652–657.

[16] S. L. Glashow, et al. *Weak Interactions with Lepton-Hadron Symmetry.* Physical Review, **2** (1970) D7, 1285–1292.

[17] J. Iliopoulos. *Glashow-Iliopoulos-Maiani mechanism.* Accessed Oct 22, 2013.

[18] PDG. *Review of particle physics.* Journal of Physics, **G37** (2010), 075021.

[19] L. Wolfenstein. *Parametrization of the Kobayashi-Maskawa Matrix.* Physical Review Letter, **51** (1983) 21, 1945–1947.

[20] B. Pontecorvo. *Neutrino Experiments and the Problem of Conservation of Leptonic Charge.* Journal of Experimental and Theoretical Physics, **26** (1968), 984–988.

[21] Z. Maki, et al. *Remarks on the unified model of elementary particles.* Progress of Theoretical Physics, **28** (1962) 5, 870–880.

[22] F. Zwicky. *Spectral displacement of extra galactic nebulae.* Helvetica Physica Acta, **6** (1933), 110–127.

[23] M. Roberts et al. *Comparison of rotation curves of different galaxy types.* Astronomy and Astrophysics, **26** (1973), 483–485.

[24] WMAP Collaboration. *Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results.* arXiv:1212.5226, (2012).

[25] W. Altmannshofer, et al. *Model-Independent Constraints on New Physics in* $\mathsf{b} \to \mathsf{s}$ *Transitions.* Journal of High Energy Physics, **1204** (2012), 008.

[26] W. Altmannshofer et al. *Cornering New Physics in* $\mathsf{b} \to \mathsf{s}$ *Transitions.* Journal of High Energy Physics, **1208** (2012), 121.

[27] F. Beaujean, et al. *Bayesian Fit of Exclusive* $\mathsf{b} \to \mathsf{s}\ell^+\ell^-$ *Decays: The Standard Model Operator Basis.* Journal of High Energy Physics, **1208** (2012), 030.

[28] N. Kosnik. *Model independent constraints on leptoquarks from* $qb \to \mathsf{s}\ell^+\ell^-$ *processes.* Physical Review, **D86** (2012), 055004.

[29] F. Beaujean, et al. *Implications of the experimental results on rare* $\mathsf{b} \to \mathsf{s}\ell^+\ell^-$ *decays.* (2013).

[30] S. Descotes-Genon, et al. *Exploring New Physics in the C7-C7' plane.* Journal of High Energy Physics, **1106** (2011), 099.

[31] S. Descotes-Genon, et al. *Understanding the* $\mathsf{B}^0 \to \mathsf{K}^{*0}\mu^+\mu^-$ *Anomaly.* Physical Review, **D88** (2013), 074002.

[32] A. J. Buras. *Weak Hamiltonian, CP violation and rare decays.* (1998), 281–539.

[33] K. G. Wilson. *Nonlagrangian models of current algebra.* Physical Review, **179** (1969), 1499–1512.

[34] BaBar Collaboration. *The BABAR physics book: Physics at an asymmetric* B *factory.* (1998).

[35] A. Pich. *Effective field theory: Course.* (1998), 949–1049.

[36] U. Nierste. *Three Lectures on Meson Mixing and CKM phenomenology.* (2009), 1–38.

[37] I. Dunietz, et al. *In pursuit of new physics with* B_s^0 *decays.* Physical Review, **D63** (2001), 114015.

[38] LHCb Collaboration. *Opposite-side flavour tagging of B mesons at the LHCb experiment.* European Physical Journal, **C72** (2012), 2022.

[39] K. D. Bruyn, et al. *Branching Ratio Measurements of* B_s^0 *Decays.* Physical Review, **D86** (2012), 014027.

[40] K. D. Bruyn, et al. *Probing New Physics via the* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *Effective Lifetime.* Physical Review Letter, **109** (2012), 041801.

[41] A. J. Buras, et al. *Probing New Physics with the* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *Time-Dependent Rate.* Journal of High Energy Physics, **1307** (2013), 77.

[42] O. L. for HFAG. private communication.

[43] LHCb Collaboration. *Measurement of the effective* $\mathsf{B}_\mathsf{s}^0 \to \mathsf{K}^+\mathsf{K}^-$ *lifetime.* Physics Letter, **B707** (2012), 349–356.

[44] É. Maurice. *Mesure de la violation de CP dans les désintégrations* $\mathsf{B}_\mathsf{s}^0 \to \mathsf{J}/\psi\phi$*, auprès du détecteur LHCb.* Ph.D. thesis, Aix-Marseille U. (2012). CERN-THESIS-2012-089, Presented 18 Jul 2012.

[45] A. J. Buras. *Relations between* $\Delta M_{s,d}$ *and* $\mathsf{B}_{(\mathsf{s})}^0 \to \mu^+\mu^-$ *in models with minimal flavor violation.* Physics Letter, **B566** (2003), 115–119.

[46] R. Dowdall, et al. B-*meson decay constants from improved lattice NRQCD and physical* u, d, s *and* c *sea quarks.* Physical Review Letter, **110** (2013), 222003.

[47] C. Davies. *Standard Model Heavy Flavor physics on the Lattice.* PoS, **LATTICE2011** (2011), 019.

[48] A. Buras, et al. *On the Standard Model prediction for* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$. European Physical Journal, **C72** (2012), 2172.

[49] C. Bobeth, et al. *QCD corrections to* $\mathsf{B}_{(\mathsf{s})}^0 \to \mathsf{X}_{\mathsf{d},\mathsf{s}}\nu\nu$, $\mathsf{B}_{(\mathsf{s})}^0 \to \ell^+\ell^-$, $\mathsf{K}_\mathsf{L}\mu^+\mu^- \to \mathsf{MSSM}$. Nuclear Physics, **B630** (2002) 1–2, 87–131.

[50] G. Buchalla, et al. *Weak decays beyond leading logarithms.* Reviews of Modern Physics, **68** (1996), 1125–1144.

[51] G. Buchalla et al. *Two loop large m(t) electroweak corrections to* $\mathsf{K}^+ \to \pi^+\nu\bar{\nu}$ *for arbitrary Higgs boson mass.* Physical Review, **D57** (1998), 216–223.

[52] J. Brod, et al. *Two-Loop Electroweak Corrections for the* $\mathsf{K} \to \pi\nu\bar{\nu}$ *Decays.* Physical Review, **D83** (2011), 034030.

[53] CDF Collaboration. *Search for* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *and* $\mathsf{B}^0 \to \mu^+\mu^-$ *Decays with CDF II.* Physical Review Letter, **107** (2011), 191801.

[54] LHCb Collaboration. *Strong constraints on the rare decays* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *and* $\mathsf{B}^0 \to \mu^+\mu^-$. Physical Review Letter, **108** (2012), 231801.

[55] CMS Collaboration. *Search for* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *and* $\mathsf{B}^0 \to \mu^+\mu^-$ *decays in pp collisions at 7 TeV.* Physical Review Letter, **107** (2011), 191802.

[56] G. Aad et al. *Search for the decay* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *with the ATLAS detector.* Physics Letter, **B713** (2012), 387–407.

[57] A. J. Buras. *Minimal flavor violation.* Acta Physica Polonica, **B34** (2003), 5615–5668.

[58] G. D'Ambrosio, et al. *Minimal flavor violation: An Effective field theory approach.* Nuclear Physics, **B645** (2002), 155–187.

[59] D. M. Straub. *Overview of Constraints on New Physics in Rare B Decays.* ArXiv:1205.6094, (2012).

[60] M. Blanke, et al. *Rare* K *and* B *Decays in a Warped Extra Dimension with Custodial Protection.* Journal of High Energy Physics, **0903** (2009), 108.

[61] K. Agashe et al. *Supersymmetric flavor models and the* $\mathsf{B} \to \phi\mathsf{K}^0_\mathsf{S}$ *anomaly.* Physical Review, **D68** (2003), 035017.

[62] G. G. Ross, et al. *Spontaneous CP violation and nonAbelian family symmetry in SUSY.* Nuclear Physics, **B692** (2004), 50–82.

[63] S. Antusch, et al. *Solving the SUSY Flavour and CP Problems with SU(3) Family Symmetry.* Journal of High Energy Physics, **0806** (2008), 068.

[64] L. J. Hall et al. *A Geometry of the generations.* Physical Review Letter, **75** (1995), 3985–3988.

[65] H. E. Logan et al. $\mathsf{B}^0_{(\mathsf{s})} \to \ell^+\ell^-$ *in a two Higgs doublet model.* Nuclear Physics, **B586** (2000), 39–55.

[66] W. Skiba et al. $\mathsf{B}^0_\mathsf{s} \to \tau^+\tau^-$ *decay in a two Higgs doublet model.* Nuclear Physics, **B404** (1993), 3–19.

[67] ALEPH. *Determination of the Number of Light Neutrino Species.* Physics Letter, **B231** (1989), 519.

[68] E. Kuflik, et al. *Implications of Higgs Searches on the Four Generation Standard Model.* (2012).

[69] A. Djouadi et al. *Sealing the fate of a fourth generation of fermions.* Physics Letter, **B715** (2012), 310–314.

[70] W.-S. Hou, et al. *Implication of possible observation of enhanced* $\mathsf{B}^0 \to \mu^+\mu^-$ *decay.* Physical Review, **87** (2013) 9, 094005.

[71] A. Leike. *The Phenomenology of extra neutral gauge bosons.* Physics Reports, **317** (1999), 143–250.

[72] T. G. Rizzo. Z' *phenomenology and the LHC.* hep-ph/0610104, (2006), 537–575.

[73] A. J. Buras, et al. *The Anatomy of Z' and Z with Flavour Changing Neutral Currents in the Flavour Precision Era.* Journal of High Energy Physics, **1302** (2013), 116.

[74] A. Dighe et al. *How large can the branching ratio of* $\mathsf{B}_\mathsf{s}^0 \to \tau^+\tau^-$ *be ?* Physical Review, **D86** (2012), 054023.

[75] S. Sahoo, et al. *Effect of both* Z^0 *and* Z'*-mediated flavor-changing neutral currents on* $\mathsf{B}_\mathsf{s}^0 \to \tau^+\tau^-$ *decay.* International Journal of Modern Physics, **A27** (2012), 1250184.

[76] V. M. Abazov et al. *Measurement of the anomalous like-sign dimuon charge asymmetry with 9* fb^{-1} *of* $\mathsf{p}\bar{\mathsf{p}}$ *collisions.* Physical Review, **D84** (2011), 052007.

[77] V. M. Abazov et al. *Evidence for an anomalous like-sign dimuon charge asymmetry.* Physical Review Letter, **105** (2010), 081801.

[78] V. M. Abazov et al. *Evidence for an anomalous like-sign dimuon charge asymmetry.* Physical Review, **D82** (2010), 032001.

[79] LHCb Collaboration. *Measurement of the flavour-specific CP violating asymmetry* a_sl^s *in* B_s^0 *decays.* LHCb-CONF-2012-022 (2012).

[80] R. Giles et al. *Two-Body Decays of* B *Mesons.* Physical Review, **D30** (1984), 2279.

Chapter 2

The LHCb Experiment

In 2010, while the two B-factories, BaBar and Belle, were shut down and the Tevatron was recording its last $p\bar{p}$ collisions, the very first pairs on protons were colliding at the LHC ring point 8 where LHCb, the B-physics dedicated experiment, is installed. This facility was designed around a physics program built on several flagship measurements among which the branching fractions of the $B^0_{(s)} \to \mu^+\mu^-$ processes. This chapter describes how the LHCb detector allows to perform such searches in the LHC hadronic environment. Section 2.1 describes the LHC environment from the point of view of B-physics. Based on these considerations, Section 2.2 describes the main characteristics of the LHCb detector.

2.1 The LHC Environment

This section describes the LHC environment from the point of view of B-physics. First, Section 2.1.1 explains how the protons are accelerated by the CERN accelerator complex. Then Section 2.1.2 highlights the accelerator settings which are essential for a B physics program. The crux of the matter, for such a program, is to produce as many clean collisions with B-mesons as possible. We will show that this number of collisions is the product of two quantities, the integrated luminosity and the production cross section. These two quantities are sensitive to different settings of the accelerator.

2.1.1 The CERN Accelerator Complex

The LHC is the final and most powerful element of the CERN particle accelerator complex settled near Geneva at the boundary between France and Switzerland. The construction of such a facility is part of a vast physics

program aiming to discover physics beyond the standard model. The key elements of this exploratory program are the energy accessible at the collision points and the collision rate. The LHC being installed in the LEP tunnel, only hadron collisions allow to reach energies above the TeV and, as explained in the next section, protons are the most suitable hadrons to achieve high luminosities. The LHC is therefore a proton proton collider. The LHC also allows to collide heavy ions in order to study the quark gluon plasma.

Protons to be accelerated and collided are obtained from hydrogen by stripping off electrons. Then a first series of devices (Radio Frequency Quadrupole or QRF, Linear Accelerator or LINAC2, Proton Synchrotron Booster or PSB) focus the protons beam and accelerate them up to 1.4 GeV. Then the Proton Synchrotron (PS) accelerates the protons up to 25 GeV and provides bunch packets with 25 ns spacing to the Super Proton Synchrotron (SPS) which raises the protons energy up to 450 GeV. The bunches are then injected into the two LHC rings in opposite directions. Once the rings are filled, the protons in each beams are ramped up to 4 TeV and beams are shaped to be collided at each LHC collision point. The full accelerator chain is shown in Figure 2.1. For a more complete description of the LHC see [1] the abriged version of the LHC Design Report [2].

Proton proton collisions, as well as other diverse mechanisms, like inelastic scattering due to collisions between protons and gas molecules remaining in the beam pipes, reduce progressively the beam proton population. When the beam current is too low, approximately ten hours after the first collisions, beams are dumped into an absorber located at point 6.

2.1.2 B Physics at the LHC Environment

For a long time, precision measurement of B-physics observables were the realm of e^+e^- colliders. However the Tevatron B-physics program [3] proved such precise measurements to be possible in a hadronic environment and, even more, complementary to the ones performed at B-factories. Indeed the two environments represent two different compromises between B meson production and B signal purity. B-factories were run mostly in conditions in which only B mesons of a given mass (B^0 and B^+) could be produced. Hence only B decay products had to be reconstructed. On the contrary, hadronic collisions allow to produce copious samples of all types of B hadrons, but their reconstruction is challenging due to the very busy underlying event. Taking the most of the hadronic environment is an interplay between the collider running conditions and the detector design. This section describes the main characteristics of LHC pp collisions at stake for an LHC based

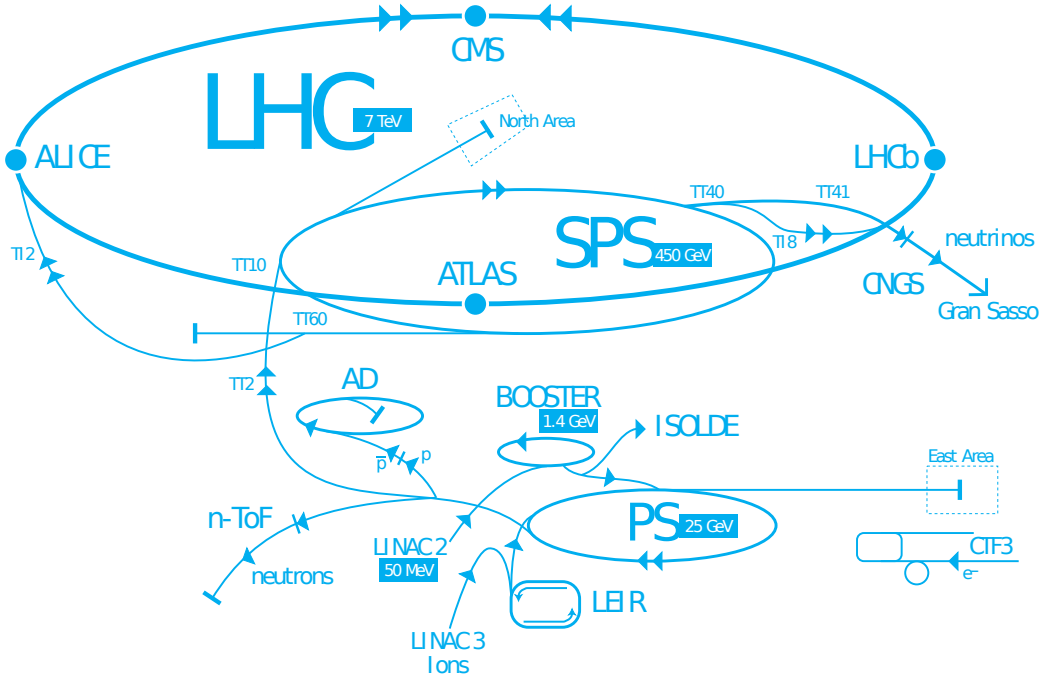


Figure 2.1: CERN accelerator complex in its nominal design. In 2011 (2012) the LHC accelerated protons only to 3.5 (4.0) TeV.

B-physics experiment.

For a physics program, the relevant quantity in which the collider running condition enters is the rate at which events are produced. This quantity can be written as:

$$N = \sigma\mathcal{L}, \tag{2.1}$$

where σ is the production cross section of the process being considered and $\mathcal{L}$ the instantaneous luminosity delivered by the collider. As far as B physics is concerned, the most interesting process is the production of b quarks, always created in pairs, $\mathrm{b\overline{b}}$. The production cross-section aspects linked to the accelerator running conditions are developed at the end of this section, as $\mathrm{b\overline{b}}$ production cross-section as well has a strong impact on the detector layout, described in the Section 2.2.

Instantaneous Luminosity

To highlight the collider properties which drive the events production, the instantaneous luminosity can be written as [4, 5]:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev}}{4\pi\sigma_x\sigma_y} F, \tag{2.2}$$

where n_b is the number of bunches, N_b the number of particle per bunch, f_{rev} the revolution frequency, $\sigma_{x,y}$ the size of the bunch in the plane transverse to the beam axis and F a geometric factor accounting for the luminosity reduction due to the crossing angle and beam separation at the interaction point (IP). The more particles circulate in the beams the higher the luminosity, therefore, since producing protons in large quantity is far easier than anti-protons, it was chosen to collide pp pairs at the LHC instead of $\mathrm{p\overline{p}}$, as at the Tevatron.

To understand the compromise at stake to reach the best running conditions for the physics, Equation 2.2 must be contrasted with Equation 2.3 which expresses the instantaneous luminosity as a function of the average number of pp interaction per bunch crossing μ [6]:

$$\mathcal{L} = \frac{\mu n_b f_{rev}}{\sigma_{ine}}, \tag{2.3}$$

where σ_{ine} is pp inelastic cross section (73.5 ± 3.1 mb for 7 TeV collisions [7]). Unlike in ATLAS or CMS, in LHCb, the detector readout rate is limited at 1 MHz. Matching this constraint with trigger thresholds allowing to select physically interesting events, requires to keep μ around 1.7. Moreover, these

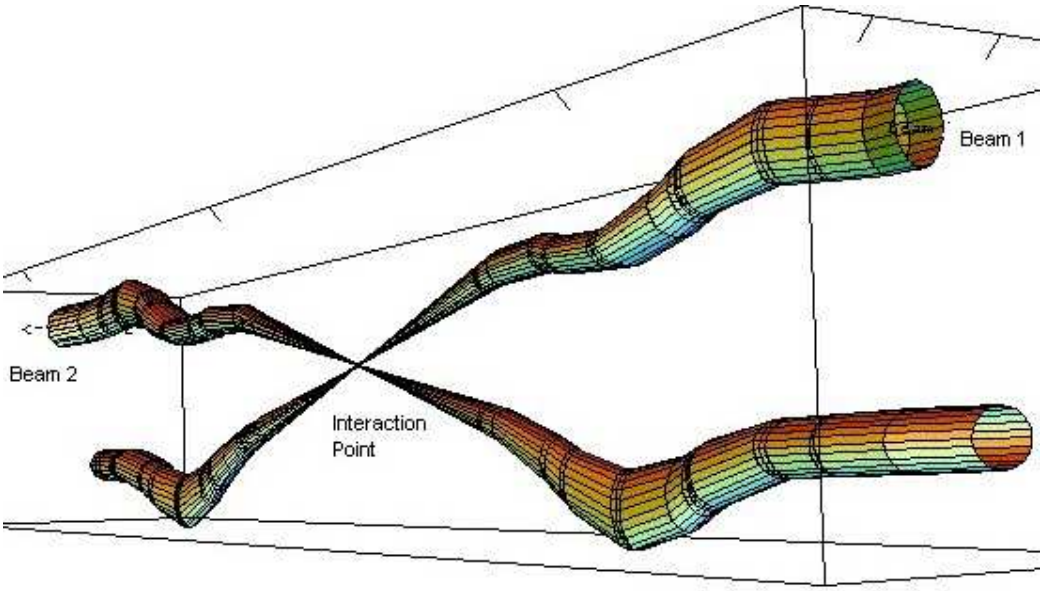


Figure 2.2: Two proton beams around an interaction point.

low μ running conditions reduce the radiation level to which the sub-detectors are exposed. Different running conditions are achieved for each experiment by mainly tuning the size of the beam ($\sigma_{x,y}$ in Equation 2.2) at its interaction point as shown in Figure 2.2.

As explained in the previous section, the beams protons population reduces progressively. At LHCb, the protons population decrease is compensated by bringing the beam closer from each other, which increases the value of the geometric factor F. This technique is called luminosity levelling [8] and allowed in 2012 to record 97% of the data with an instantaneous luminosity within 3% of the 2012 nominal one. Such a stability of the running conditions allows to reduce and better control systematic uncertainties. At ATLAS and CMS on the contrary, the geometric factor being already at its maximum, the protons population decrease translates into an exponential decay of the instantaneous luminosity, as shown in Figure 2.3.

The last aspect of the collider running condition which has a strong impact on the production mechanisms is the beam energy. Indeed as shown in Figure 2.4 (taken from [9]), the production cross sections of all processes depend on the centre of mass energy. The higher the energy at the centre of mass, the larger the $\mathsf{b\bar{b}}$ cross section and the higher the ratio of this cross section to the total one. Figure 2.4 shows also that, up to a good approximation, the $\mathsf{b\bar{b}}$ cross section increases linearly with the centre of mass energy. This

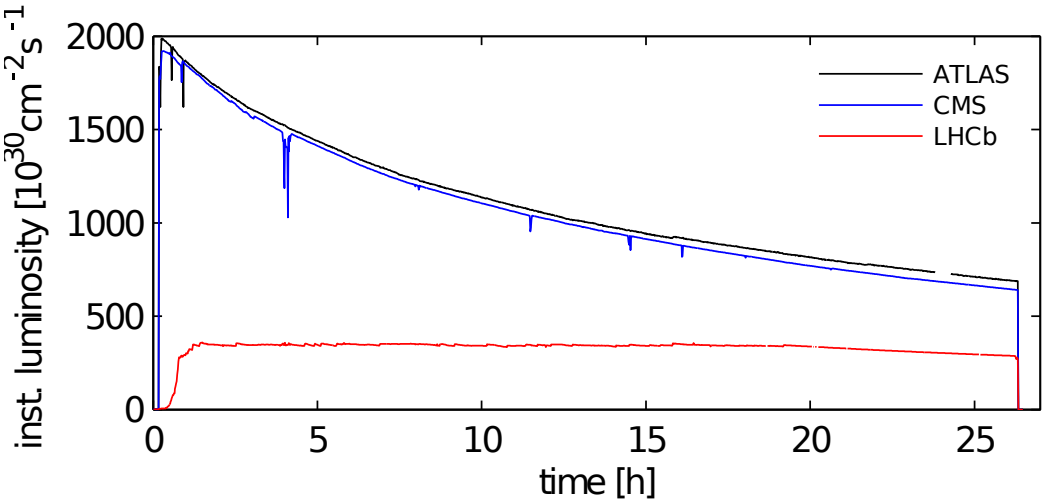


Figure 2.3: Example of instantaneous luminosities at ATLAS, CMS, and LHCb during fill number 2006.

behaviour will be used to project the sensitivity of the $B_s^0 \to \mu^+\mu^-$ analysis (see Section 3.3.4) and in particular to account for the raise of 1 TeV in the centre of mass energy between 2011 and 2012. At 7 TeV the $b\bar{b}$ production cross section is $284 \pm 53\,\mu$b, hence for a luminosity of $4 \times 10^{32}\,\text{cm}^{-2}\text{s}^{-1}$, around 4×10^{11} $b\bar{b}$ pairs are produced each $1\,\text{fb}^{-1}$ ($1.0\,\text{fb}^{-1}$ were recorded in LHCb in 2011 and $2.0\,\text{fb}^{-1}$ in 2012).

$b\bar{b}$ Production Cross Section

The $b\bar{b}$ production cross section at the LHC features in fact a non trivial azimuthal angle distribution. As shown in Figure 2.5 the $b\bar{b}$ pairs are mainly produced in the forward region. The processes responsible for this production, illustrated in Figure 2.6, are as follows [10, 11, 12]:

flavour excitation (54% of the production): a $b\bar{b}$ pair from the proton quark sea is excited into the final state due to one of the b quarks undergoing a hard QCD interaction with a parton from the other proton,

gluon splitting (27% of the production): a $b\bar{b}$ pair arises from a $g \to b\bar{b}$ splitting at the hard scattering initial or final state,

pair production (16% of the production): a $b\bar{b}$ pair arises from quarks or gluons fusion, $q\bar{q} \to b\bar{b}$ or $gg \to b\bar{b}$,

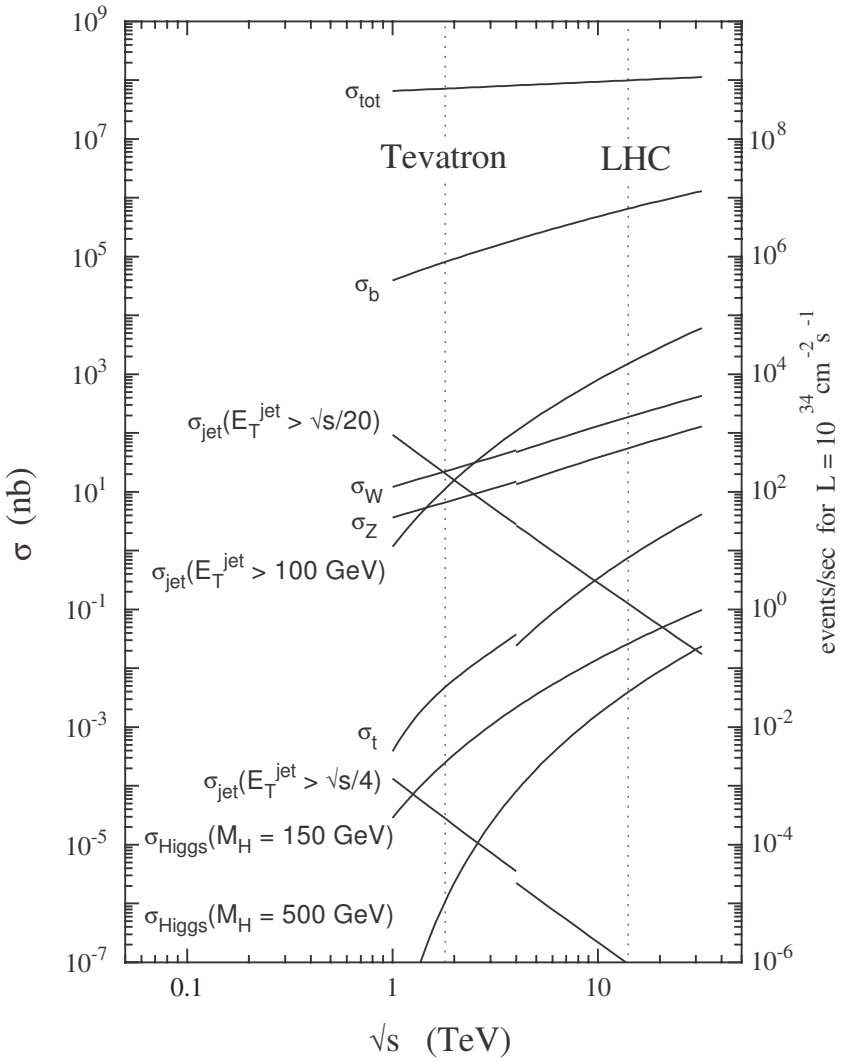


Figure 2.4: Production cross section of various processes as a function of the pp centre of mass energy.

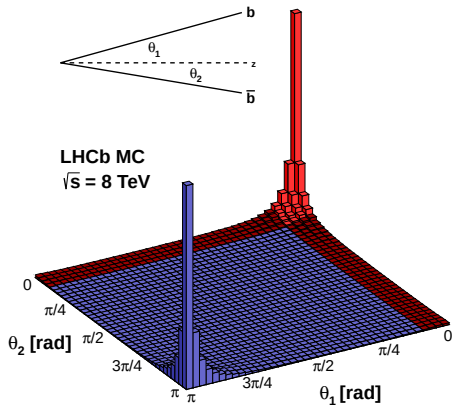


Figure 2.5: Angular correlation between the b and $\overline{\mathsf{b}}$ produced at the LHC according to simulated data [11, 12].

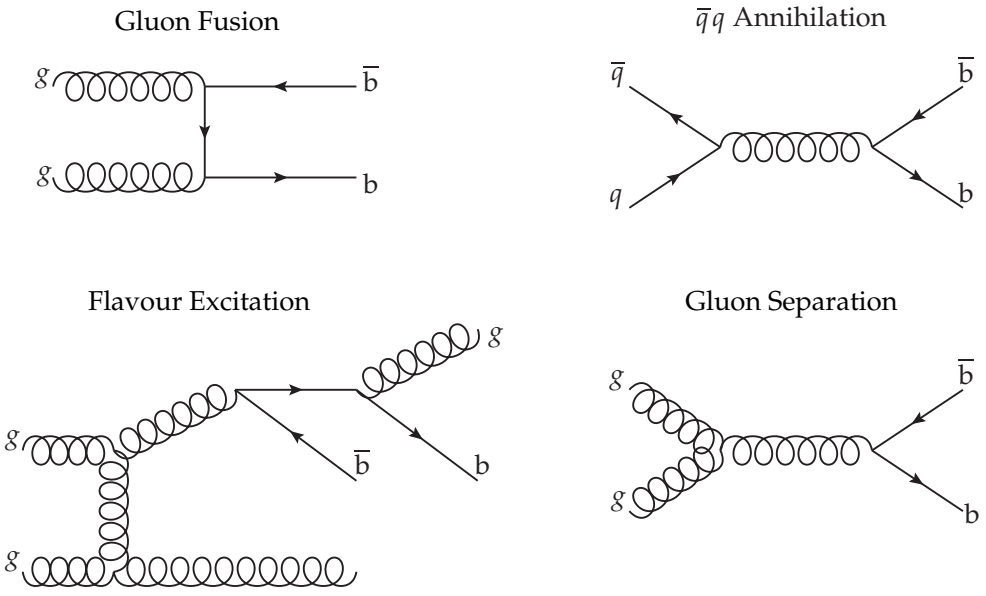


Figure 2.6: $\mathsf{b}\overline{\mathsf{b}}$ production mechanisms at LHCb.

2.2 The LHCb Experiment

The description of the physical conditions at the LHC, held in Section 2.1, outlines the main characteristics a B-physics dedicated detector must have to take the most out of this environment. This section describes the detector designed by the LHCb collaboration. First a general description is given where the main detector functions are identified. Then the sub-detectors fulfilling each functions are briefly described with a few words on their performances. For a review of the detector technologies mentioned in the section, the reader is referred to Section 31 of [13].

2.2.1 General Description

As mentioned in the previous section, the azimuthal angular distribution of the $b\bar{b}$ pairs produced at the LHC is peaked in the forward regions. For this reason the LHCb was designed to instrument the interaction point in the forward region. Its acceptance comprises all polar angles, θ, between 10 and 250 mrad with respect to the beam axis which corresponds to pseudo-rapidities,

$$\eta \;=\; -\ln(\tan\frac{\theta}{2}), \tag{2.4}$$

between 2 and 5. Figure 2.7 shows the LHCb geometrical coverage and its complementarity to ATLAS and CMS.

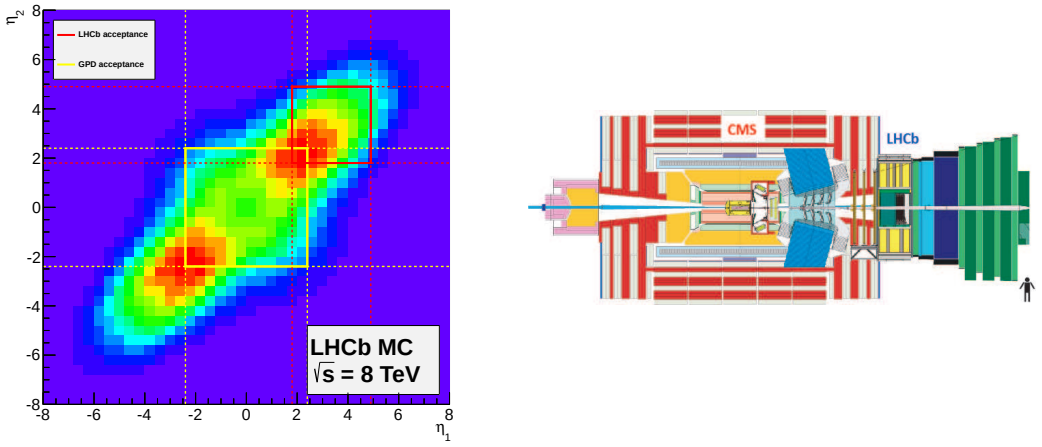
The instrumentation installed in the forward region aims to three main purposes:

- reconstruct particles' momenta and tracks,
- identify particles,
- trigger interesting events.

Each of these purposes is achieved with a set of dedicated devices, briefly described in the following section. The main constraint for the detectors design are the minimisation of the material budget, the radiation damage resistance, and the readout time. The global layout of the detector is shown in Figure 2.8. For a more detailed description, see [14].

2.2.2 Tracking

The tracking system is essential to reconstruct a charged particle as it provides its track, charge, and momentum. Track reconstruction is of particular



(a) (b)

Figure 2.7: (a) Correlation between the b and $\overline{\mathsf{b}}$ pseudo-rapidity produced at the LHC. The red (yellow) box represents the LHCb (ATLAS and CMS) coverage. (b) Comparison between the CMS and LHCb acceptances.

interest for this thesis as tracks are the starting point of the inclusive topological based vertex reconstruction algorithm used for the $\mathsf{B}^0_\mathsf{s} \to \tau^+\tau^-$ analysis.

The tracking system is composed of the *Vertex Locator* (VELO) [15], which provides precise measurements of tracks coordinates close to the interaction point, a dipole *magnet* [16], which bends the tracks according to their charges and momenta, situated between planar tracking station, namely the *Tracker Turicensis* (TT), upstream of the magnet, and the *T1-T3* station downstream. The T1-T3 planar tracking stations are implemented with two technologies: silicon trackers in the region close to the beam pipe [17] and drift chambers for the outer region [18], as the required granularity is lower, due to the smaller track density. All tracking systems cover the full LHCb acceptance.

Vertex Locator

The position of the B meson decay is a crucial information for the trigger system and for many physics analysis, either to disentangle the B signal candidates from short lived backgrounds or to perform time dependent analyses. In LHCb, the B mesons are boosted and fly around 1 cm away from the primary vertex before decaying. To achieve precise measurements of these vertex positions, the VELO sensors are placed 8 mm from the collision point.

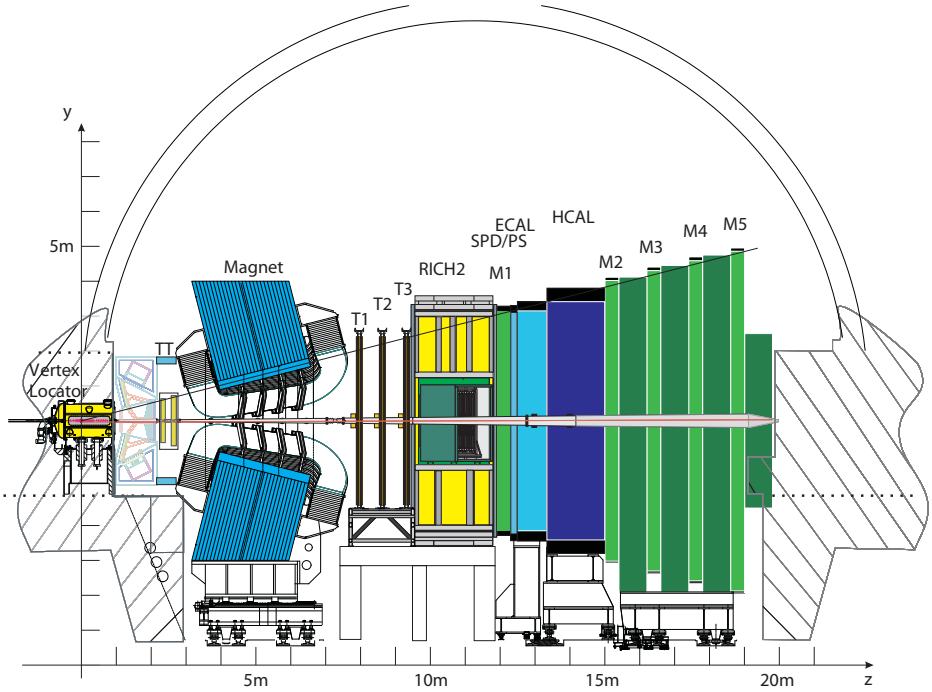
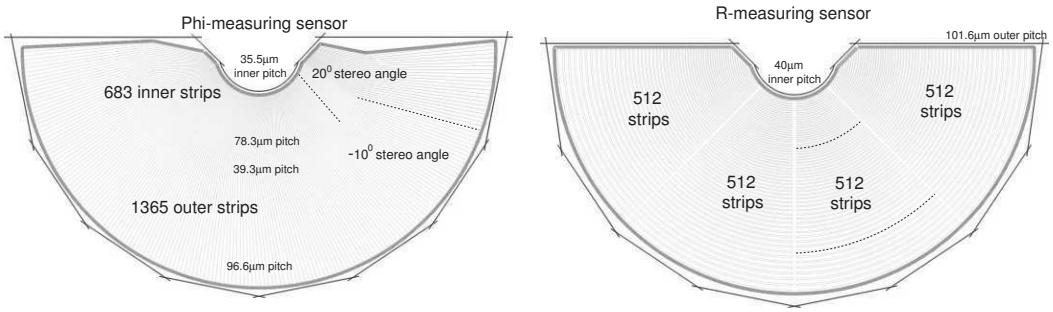


Figure 2.8: Side view of the LHCb detector.

This distance being smaller than the LHC aperture during injection, the VELO has the ability to retract and closes only when beams are stable.

The VELO sensors are silicon modules of two kind of geometry, one segmented in ϕ and the other in r, as shown in Figure 2.9. This layout was chosen as it allows a fast and efficient determination of the track impact parameter for the high level trigger.



(a) (b)

Figure 2.9: ϕ(a) and r (b) VELO sensors.

These semicircle sensors are arranged into two 1 meter long semi-cylinder of twenty-one $r-\phi$ sensors pairs, with a spacing along the beam axis increasing with the distance to the interaction point, as shown in Figure 2.10.

Two additional ϕ sensors placed before the interaction point shown in Figure 2.10 are used to measure the number of pp interactions per crossing and provide pile-up information to the trigger.

An idea of the VELO performances is given in Figure 2.11, which shows the primary vertex resolution along the beam axis (78.3 μm with twenty five tracks events).

Magnet

The magnet is used to measure the momenta of charged particles. It is a dipole magnet with an integrated field of 4 Tm for 10 m long tracks which allows to achieve a momentum precision at 5 GeV of $\delta p/p = 0.4\%$. Such a performance requires also to know precisely the field map which is therefore measured with a array of Hall probes spanning from the interaction point to the RICH2 detector. Finally, to cancel the systematic uncertainties for CP asymmetries measurements, the magnet polarity can be reversed periodically.

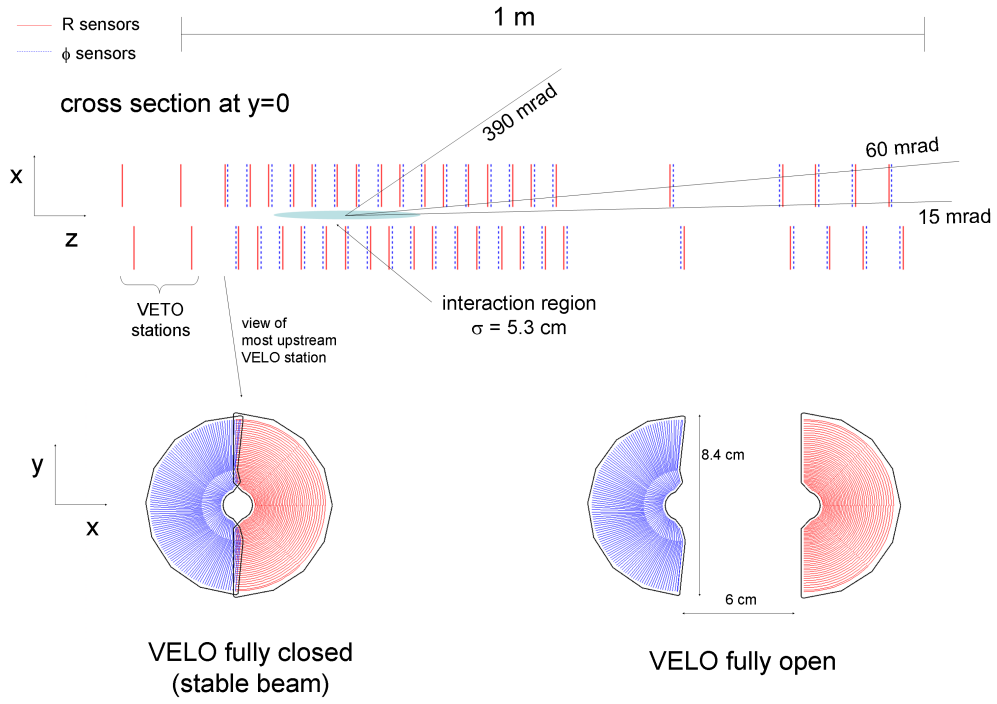


Figure 2.10: VELO dectector layout.

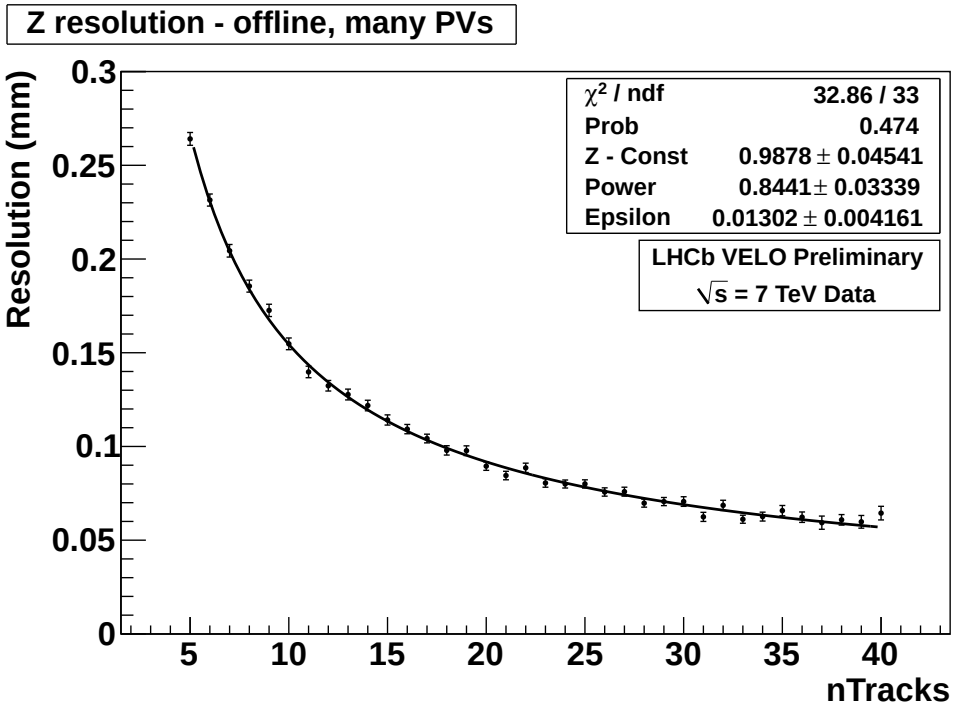
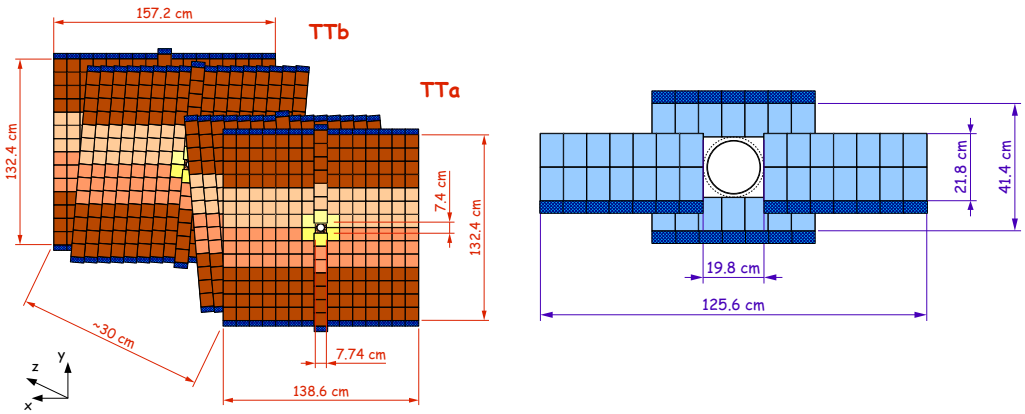


Figure 2.11: Primary vertex resolution along the beam axis measured in 2011 data by randomly splitting the track sample in two and comparing measured positions. Events used have more than one reconstructed primary vertex.

Silicon Trackers (TT and IT1-3)

The TT and the inner part of the T1-T3 tracker stations consist of four detection layers ($xuvy$) with vertical 200 μm wide strips in the first and last layers (x and y) and strips rotated by a stereo angle of $-5°$ and $+5°$ in the second and third layer (u and v), as shown in Figure 2.12. This layout allows to reach 50 μm resolution on the horizontal hit position and represents a compromise between the tracking algorithms efficiency, the combinatorics tracks rate, and the computing time (see Section 4.5 of [18] for a more detailed description of the angle choice).



(a) (b)

Figure 2.12: Layout of the $xuvy$ TT layers(a) and of one x IT layer (b).

Outer Trackers (OT1-3)

Silicon technology is precise but expensive and such a granularity is only needed in the inner region of the detector where the track density is high, to avoid high detector occupancy. Therefore the vast ($5971 \times 4850\,\mathrm{mm}^2$) outer part of the T1-T3 station, shown in Figure 2.13, can be equipped with coarser granularity detectors. These detectors are four layers of straw-tubes modules arranged in the same $xuvy$ geometry as the IT ones. Each module is 34 cm wide and made of two staggered layers of 64 drift tubes each. The straw tubes are cylinders filled up with argon (70%) and carbon-dioxide (30%). A high voltage tension is applied between the inner carbon-doped Kapton made part of the cylinder and a gold-plated tungsten wire sucked through the tube. This tension accelerates electrons ionized by a particle passing towards the wire. These electrons are then collected on the wire, in less than 50 ns which

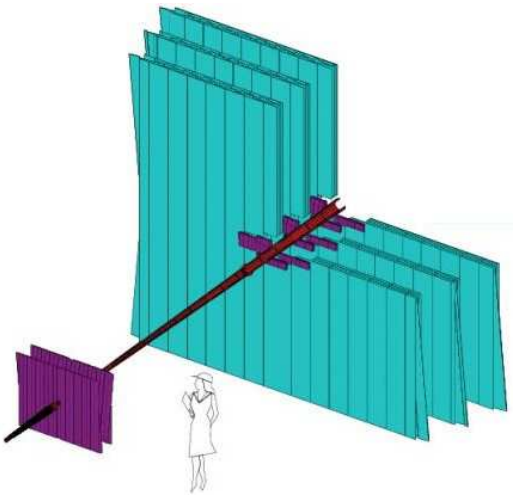


Figure 2.13: TT and T1-T3 tracking system. The purple pads are equipped with silicon technology, the cyan ones with straw tubes.

allows to reach 200 μm resolution on the radial coordinate. This process is illustrated in Figure 2.14.

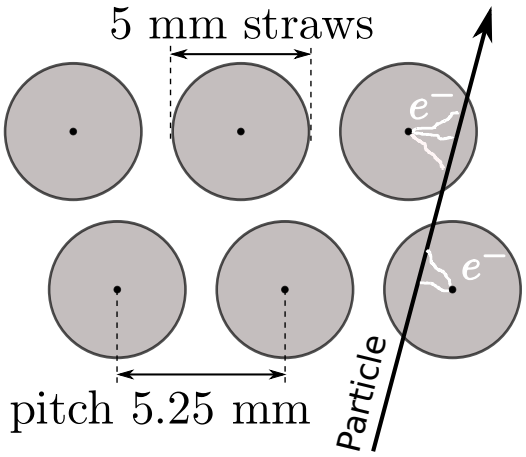


Figure 2.14: Straw tube particle detection principle in the outer tracker layout.

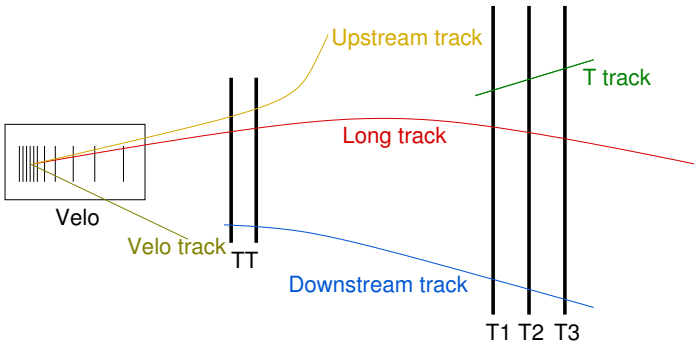


Figure 2.15: Track type returned by the different tracking strategies.

Tracking Algorithm

The trajectory of the particles are reconstructed from all the hits collected in the tracking systems. Two stages are distinguished in the reconstruction algorithm. First detector hits are associated to track candidates according to a given strategy, then the track candidates are fitted with the Kalman filter technique [19]. There are several tracking strategies to associate detector hits to candidate tracks [20]. These strategies are not independent. Indeed, the strategy returning more refined track types (e.g. LongTrack) use as input simpler track types (e.g. 3D Velo tracks). All the track types are represented in Figure 2.15. The tracks with the best momentum and impact parameter resolution are the Long Tracks, hence they are the main type of tracks used for physics analyses.

Since several tracking strategies are used, a real particle can be associated to several track candidates. These multiple candidates are called clones and are detected by running a clone killer algorithm [21] which keeps only the best of the clones. Some tracks are also created from random combinations of hits in the detectors. These tracks are called ghost. For each track the probability to be ghost is computed with a likelihood method described in [22].

The Kalman filter returns for each track the χ^2 of the fit which quantifies the track quality and a covariant matrix which allows to estimate the probability that the particle passed at a given point[1].

As far as performances are concerned, the tracking system and algorithm allow to reconstruct long tracks with a 96% efficiency.

[1]The covariant matrix is a key ingredient of the inclusive event vertex reconstruction described in Section 5.

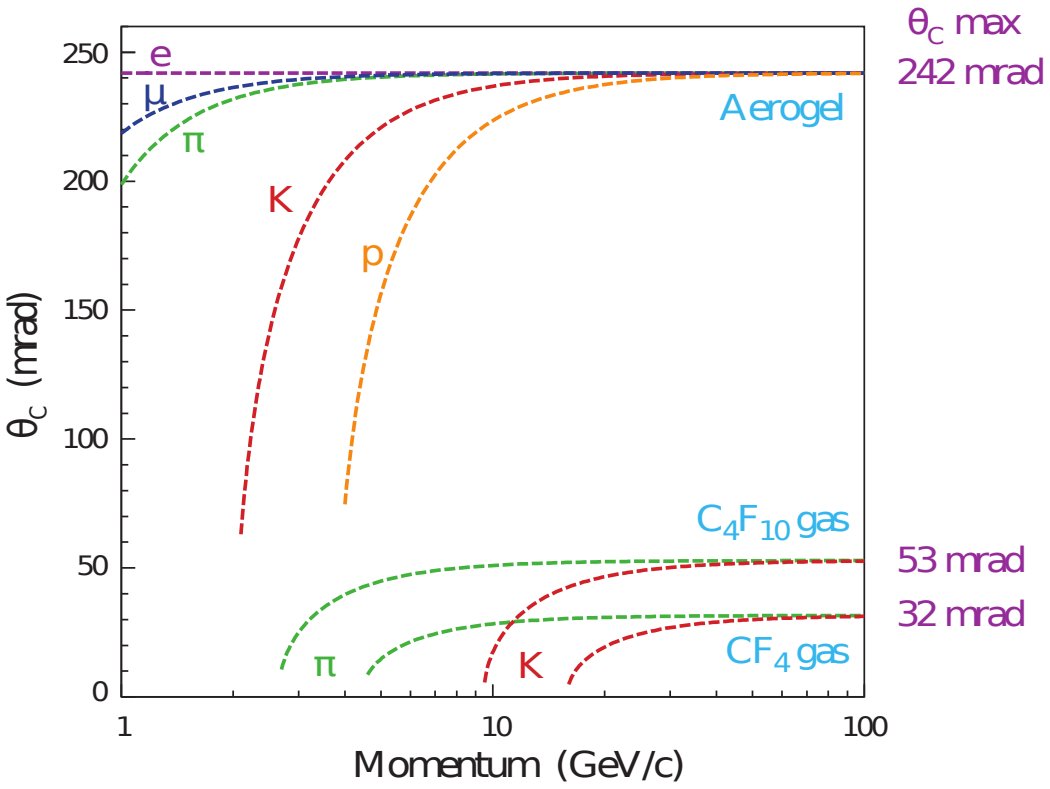


Figure 2.16: Cherenkov angle as a function of particles momentum for different particles and different media.

2.2.3 Particle Identification

RICH

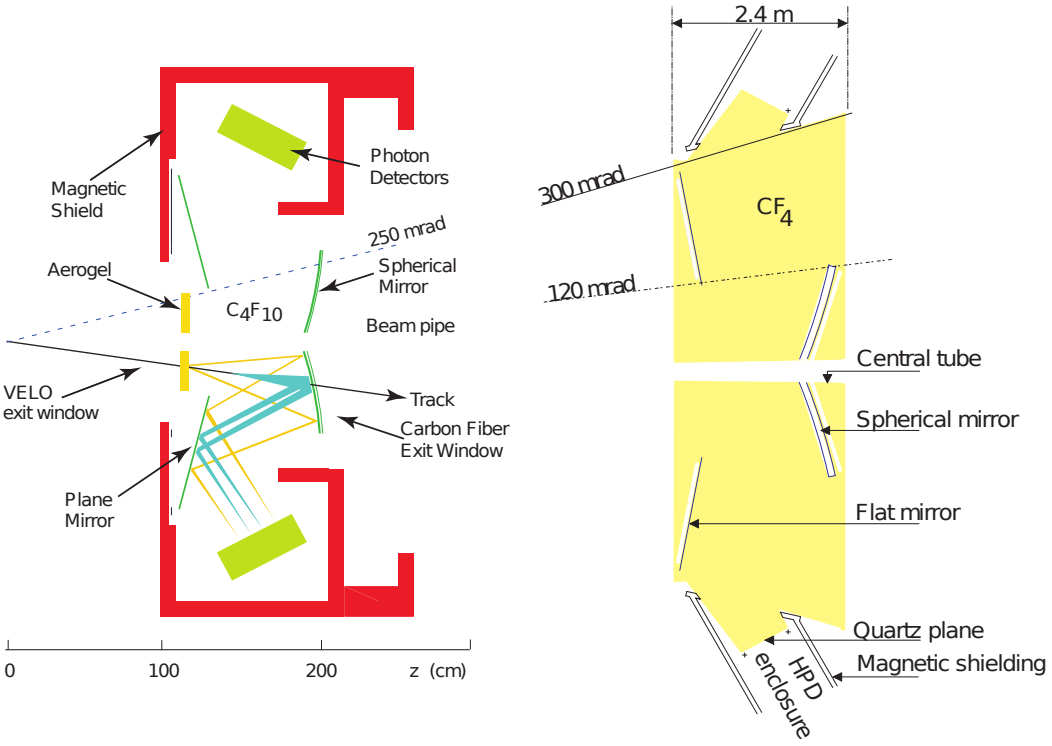
Charged particles can be identified exploiting the Cherenkov effect. When a charged particle travels at a velocity larger than the local medium velocity it radiates a light cone at an angle θ_c with respect to the particle direction which depends on the velocity. Measuring the angle of this light cone allows to access the particle speed. Knowing the particle charge and momentum, its mass can be inferred, hence the particle identified. Figure 2.16 shows for different particles and the different phases used in the LHCb RICH detectors how θ_c evolves as a function of the particle momentum. LHCb is equipped with two Cherenkov detectors with different media covering two momentum ranges. Low momentum particles ($1-60\,\text{GeV}$) are identified with the RICH1 installed upstream of the magnet and using aerogel and fluorobutane (C_4F_{10}) gas media. Higher momentum particles ($15-100\,\text{GeV}$) are identified with

the RICH2 installed downstream of the magnet and the tracking stations and using a CF_4 gas medium.

While the RICH1 covers the full LHCb acceptance, the RICH2 has an acceptance limited at ±15 mrad to ±120 mrad (horizontal) and ±100 mrad (vertical) as high momentum particles tend to be forward.

The photons emitted are then collected with an optical system of spherical and flat mirrors and detected with Hybrid Photon Detectors (HPD) which are enclosed in iron shields allowing the detectors to operate despite an external magnetic field up to 50mT. Both RICH layouts are shown in Figure 2.17.

These sub-detectors allow to reach 95% kaon identification efficiency for a 5% pion to kaon misidentification probability.



(a) (b)

Figure 2.17: Schematic view of the RICH1 (a) and RICH2 (b)

Calorimeters

The calorimeters contribute to the particle identification, as the particle energy deposit patterns in the calorimeters depend on their nature, but its main purpose is to provide fast information for the large transverse momentum

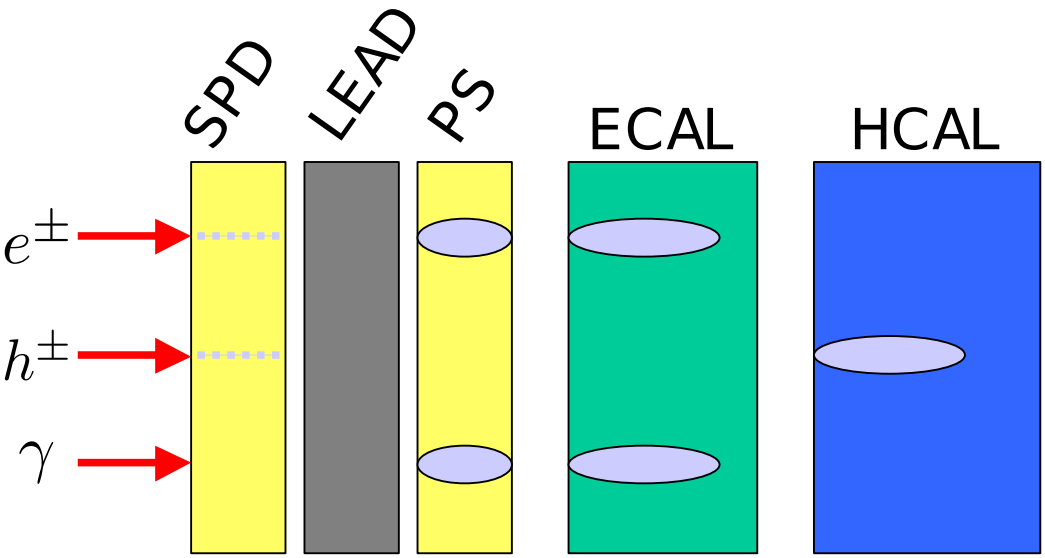


Figure 2.18: Energy deposit on the different calorimeter parts by electrons, charged hadrons, and photons. Dashed lines represent energy deposits at the minimum particle energy loss, ellipses showers.

level-0 electromagnetic and hadronic triggers. They also allow to measure photon and neutral pion energy. The calorimeters are located downstream of the RICH2 and are composed of four sub-detectors in which photons, charged hadrons and electrons depose different hit patterns from which they can be identified as shown in Figure 2.18.

The particles pass first through a scintillating pad detector (SPD) in which only charged particles depose energy. This device allows therefore to distinguish charged particles from neutral ones. Then particles pass through a 12 mm deep lead converter which is enough to initiate electrons and photons showers but too thin for the charged hadrons. These showers are measured in a second scintillating pad called pre-shower (PS) installed after the converter. The charged hadrons keep deposing in this detector and in the following one energy at their minimal energy loss. The following detector is the electromagnetic calorimeter (ECAL), composed of alternate layers of lead converters and scintillating pads. Electrons and photons continue showering in this detector which is wide enough to absorb and measure all their energy. Finally, charged hadrons initiate shower into the hadronic calorimeter (HCAL) composed of alternate layers of iron converters and scintillating pads, which is wide enough to absorb and measure all their energy.

The scintillating pads granularity is increasingly coarser going to the outer

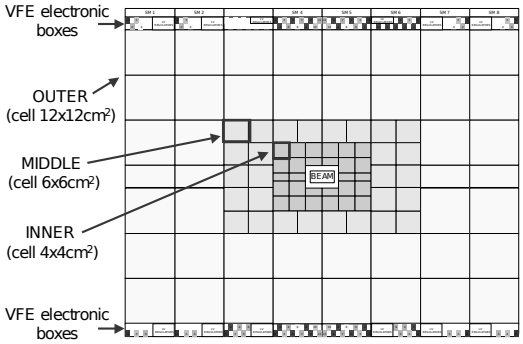


Figure 2.19: SPD, PS and ECAL detectors scheme.

region where the track multiplicity is lower as show in Figure 2.19.

The calorimeter allows to reach 90% electron identification efficiency for a 5% electron to hadron misidentification probability.

Muon System

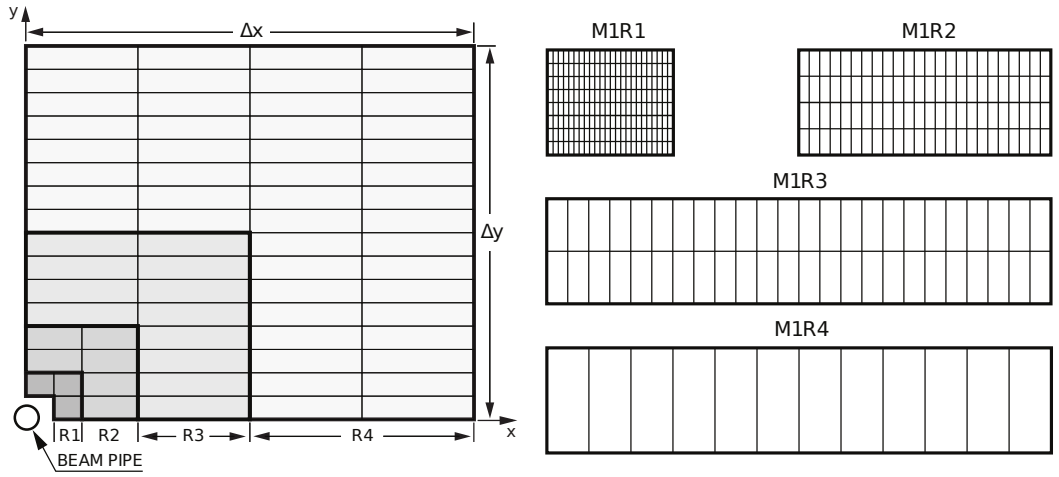
Downstream of the calorimeter the only charged particles still not absorbed are muons. Four tracking stations (M2-M4) are therefore installed in this region to provide fast information for the high transverse momentum level-0 muon trigger and particle information for the software trigger and off-line analysis. The stations are interleaved with iron absorber so that only muons with momenta larger than 5 GeV/c can penetrate through all chambers. The transverse momentum measurement for the trigger system (see Section 2.2.4) is improved by a fifth station (M1) installed in front of the calorimeter. The station granularity is increasingly coarser with the distance from the interaction point and is thinner in the bending plane as illustrated in Figure 2.20.

All chambers are Multi-wire proportional chambers (MPWC) except the inner part of the M1 which is made of of triple Gas Electron Multiplier (GEM) detectors to cope with the high particle rate.

Muon Identification

The muon identification is performed with a three step algorithm using information from both tracking and muon systems [23]. The steps of the algorithm are described in the following.

Field of Interest Definition First tracks are extrapolated linearly to the muon stations. These extrapolations allow to define, for each track, a



(a) (b)

Figure 2.20: Front view of a quadrant of a muon station (a). Each rectangle represents one chamber. Division into logical pads of four chambers belonging to the four regions of station M1 (b). In each region of stations M2-M3 (M4-M5) the number of pad columns per chamber is double (half) the number in the corresponding region of station M1, while the number of pad rows per chamber is the same.

field of interest (FoI) in the muon stations where the particle, if it was a muon, could have deposed hits.

Muon Candidate Selection Then, a track becomes a muon candidate (*isMuon* variable is set to *true*) if hits are found in its FoI in two stations among M2, M3, and M4, for tracks with momenta between 3 and 6 GeV/c, and in three stations among M2, M3, M4, and M5, for tracks with larger momenta.

Muon Likelihood Computation Finally, probabilities to be muons and non-muons are computed for each candidate by comparing the distance from the position of the track extrapolation in each of the muon stations to the closest hit (in unit of the sub-detector pad size) and the distance expected for muon and for non-muon tracks. The algorithm returns the logarithm of the ratio of these two probabilities which is indicated later as $\mathsf{DLL}(\pi, \mu)$

This algorithm allows to achieve a 97% muon identification efficiency for a $(1-3)\%$ misidentification ($\pi \rightarrow \mu$) probability.

Table 2.1: Main TCKs used in 2012.

Index	TCK	Integrated Luminosity	Magnet Polarisation
TCK1	0x008c0040	$70\,\mathrm{pb}^{-1}$	Down
TCK2	0x0094003D	$280\,\mathrm{pb}^{-1}$	Up and Down
TCK2	0x0097003D	$280\,\mathrm{pb}^{-1}$	Up and Down
TCK3	0x00990042	$620\,\mathrm{pb}^{-1}$	Up and Down

2.2.4 Trigger

Figure 2.4 shows that signal processes ($\mathrm{b\bar{b}}$ pairs) production cross section is more than two orders of magnitude smaller than the total production cross section. In others words, a substantial part of the pp collisions will not produce any interesting events. Hence, the strategy adopted by LHCb is to produce as many collisions as possible - controlling the event track multiplicity to ensure a good event reconstruction - but record only the interesting events. This strategy is very challenging for precision measurements as any event selection introduces unavoidable systematic uncertainties which need to be controlled.

The collision rate being extremely high, the event selection is performed by a two levels trigger system. First, the *Level 0 trigger* (L0) made with custom electronics and using only fast available sub-detector information to be able to operate at the LHC bunch crossing (40 MHz) performs a loose selection. This selection reduces the event rate down to a maximum of 1 MHz, of which 400 kHz are dedicated to events containing muons with interesting features. At this rate, the full sub-detectors can be read and a software trigger called High Level Trigger (HLT), running on a commercial computer farm called Event Filter Farm (EFF), performs a more refined selection. This selection reduces the rate down to a maximum rate of 3 kHz.

The whole trigger configuration is written into a unique Trigger Configuration Keys (TCK) which is a 32 bit label pointing to a database. The constant instantaneous luminosity achieved with the luminosity levelling technique, allowed the triggers rate to be very stable during the 2012 data taking and only five different TCK shown in Table 2.1 were used with minor changes between them. We will refer in the following to these TCK as TCK1, TCK2, and TCK3 as TCK `0x0094003D` and `0x0097003D` differs only by some technicalities.

Level 0 Trigger

At the L0 trigger, only simple features of B decays are looked for: large transverse momentum (p_T) or energy (E_T). This search is performed by two independent triggers called *L0-Calo* and *L0-Muon*[2] which in practice return positive decision if they find:

L0 Muon a pair of large p_T muons or one large p_T muon,

L0 Calo one large E_T hadron, or photon, or electron.

These five trigger lines are referred to as `L0SingleMuon`, `L0Dimuons`, `L0Hadron`, `L0Electron`, `L0Photons`. Only the first three lines are of interest for the analyses presented in this document and their selection criteria are shown in Table 2.2. On top of these five trigger line decisions, a common Global Event Cut (GEC) is applied to skip high multiplicity events which are longer to process by the HLT. At the low level trigger, the GEC practically requires to have a maximum number of hit in the SPD. These GEC thresholds are shown in Table 2.2.

Table 2.2: L0 trigger line requirements.

Line		Criteria	GEC
`L0SingleMuon`	TCK1	$p_T > 1.48\,\text{GeV}/c$	Nb SPD hits < 600
	TCK2-3	$p_T > 1.76\,\text{GeV}/c$	
`L0Dimuon`	TCK1	$\sqrt{p_{T1} \times p_{T2}} > 1.3\,\text{GeV}/c$	Nb SPD hits < 900
	TCK2-3	$\sqrt{p_{T1} \times p_{T2}} > 1.6\,\text{GeV}/c$	
`L0Hadron`		$E_T > 3.5\,\text{GeV}/c$	Nb SPD hits < 600

High Level Trigger

Events accepted by the L0 are sent by the front-end electronics to the EFF where they are built and placed on a buffer which is then accessed by the C++ HLT program. This program refines the selections performed by each L0 trigger line. For such a refined selection, the events need to be reconstructed but, given the L0 ouput rate and the available resources in the EFF, only $\sim 30\ \mu s$ are available to treat an event. This is not sufficient for a full event reconstruction. Therefore the HLT selection is performed in two stages of increasing complexity, called HLT1 and HLT2.

[2] A third trigger called L0-Pileup exist, but is only used for luminosity measurement [24].

A first series of HLT1 programs refine the L0 trigger line selections based on partial reconstruction, using VELO and tracking stations' information, and reduce the event rate down to 40 kHz. Then a series of HLT2 programs perform selections based on a more complete event reconstruction. These HLT1 and HLT2 trigger lines apply also some GEC to limit the processing time, but their design rely mainly on the physics to be performed with data. The trigger lines selection criteria and GEC of interest for the $\mathsf{B}^0_{(\mathsf{s})} \to \mu^+\mu^-$ analysis presented in this document are shown in Table 2.3 for HLT1 and in Table 2.5 for HLT2. They use the track χ^2 per degree of freedom (χ^2/ndf), the track impact parameter (IP) and its χ^2 (IPχ^2), the track momentum (p) and transverse momentum (p_T), the number of hits in the VELO, IT, and OT, the $isMuon$ variable defined in Section 2.2.3, and the distance between the primary and the secondary vertices divided by its error (DLS). For $\mathsf{B}^0_{(\mathsf{s})} \to \tau^+\tau^-$, the main trigger HLT lines are the `Hlt1TrackAllL0` and `HLT2Topo` [25, 26]. The selection performed by `Hlt1TrackAllL0` are reported in Table 2.4. `HLT2Topo` operates a more sophisticated selection applying a multivariate discriminator. For a description of this selection we report the reader to [26].

Before 2012, the EFF was used at the maximum of its performances during data taking, but was empty at any other time. In order to use also this available processing time, a deferred HLT triggering was devised so that non-processed events could be stored on disk during data taking and only processed afterwards.

The 2011 trigger performances are documented in [27].

Table 2.3: HLT1 trigger lines' requirements.

Line		Criteria		
`Hlt1DiMuonLowMass`	TCK1	$IP\chi^2 > 3$	$p_{T\mu} > 0.5\,\text{GeV}/c$	$m_{\mu\mu} > 1.0\,\text{GeV}/c^2$
	TCK2-3	$IP\chi^2 > 6$		
`Hlt1DiMuonHighMass`	TCK1-2-3	$p_T > 0.5\,\text{GeV}/c$	$m_{\mu\mu} > 2.5\,\text{GeV}/c^2$	
`HLT1TrackMuon`		$p_T > 1\,\text{GeV}/c$	$IP > 100\,\mu\text{m}$	$IP\chi^2 > 16$
`Hlt1TrackAllL0`		$p_T > 1.85\,\text{GeV}/c$	$IP > 100\,\mu\text{m}$	$IP\chi^2 > 16$
GEC		VELO Hits $< 10^4$	IT Hits $< 3,000$	OT Hits $< 15,000$

Table 2.4: The selection criteria applied in `Hlt1TrackAllL0`.

Track IP [mm]	> 0.1
Number VELO hits/track	> 9
Number missed VELO hits/track	< 3
Number OT+IT$\times$2 hits/track	> 16
Track IPχ^2	> 16
Track p_T[GeV/c]	> 1.7
Track p [GeV/c]	> 10
Track χ^2/ndf	< 2.5

Table 2.5: HLT2 trigger lines' requirements.

Line	Criteria		
`Hlt2DiMuonJPsi`	$\text{isMuon}_{\mu 1,2} = true$	$\lvert m_{\mu\mu} - m_{J/\psi}\rvert < 120\,\text{GeV}/c^2$	
`Hlt2DiMuonDetached`	$\text{isMuon}_{\mu 1,2} = true$	$\lvert m_{\mu\mu} - m_{J/\psi}\rvert < 120\,\text{GeV}/c^2$	DLS > 3
`Hlt2DiMuonBmm`	$\text{isMuon}_{\mu 1,2} = true$	$m_{\mu\mu} > 4.5\,\text{GeV}/c^2$	DLS > 3
GEC	VELO tracks < 350		

2.3 Conclusion

This chapter described the LHCb detector and its operating conditions. First the properties of the LHC hadronic environment, in which this detector operates, were portrayed. Some of these properties, like the instantaneous luminosity, can be tuned at the experiment interaction point in order to obtain the best conditions to perform B physics analyses. On the other hand, other properties, like the production cross section, are inherent to the pp collision and dictate to some extent the detector design, like the need for a trigger system. This design aims to achieve three main goals, the particles momenta and tracks reconstruction, their identification, and the selection of the most interesting events. All sub-detectors taking part in the realisation of these functions were briefly described with a few words on their performances.

These performances are essential to the $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ analysis. Indeed they set the signal reconstruction efficiency and fix the best achievable separation between signal and background. In practice, these performances allowed to obtain, in 2012, the first evidence of the $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ decay. The analysis leading to this achievement is described in the next chapter.

Bibliography

[1] L. Evans et al. *LHC Machine.* Journal of Instrumentation, **3** (2008) 08, S08001.

[2] O. S. Brüning, et al. *LHC Design Report.* CERN-2004-003-V-1 (2004).

[3] K. Anikeev, et al. B *physics at the Tevatron: Run II and beyond.* hep-ph/0201071, (2001).

[4] W. Herr et al. *Concept of luminosity.* Proceedings CAS2003, CERN-2006-002, (2006).

[5] H. Burkhardt et al. *Absolute Luminosity from Machine Parameters.* LHC-PROJECT-Report-1019. CERN-LHC-PROJECT-Report-1019 (2007).

[6] G. Aad, et al. *Luminosity determination in pp collisions at* $\sqrt{s}$*= 7 TeV using the ATLAS detector at the LHC.* European Physical Journal, **C71** (2011) 4, 1–37.

[7] G. Antchev et al. *First measurement of the total proton-proton cross-section at the LHC energy of* $\sqrt{s} = 7 TeV$. European Physics Letter, **96** (2011) 2, 21002.

[8] G. Papotti, et al. *Experience with offset collisions in the LHC.* TUPZ025, Proc. IPAC, (2011).

[9] P. Jenni, et al. *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report.* Number ATLAS-TDR-016 in Technical Design Report ATLAS (2003).

[10] V. Andreev. *Heavy Quark Production at CMS and ATLAS.* (2009).

[11] T. Sjostrand, et al. *PYTHIA 6.4 Physics and Manual.* Journal of High Energy Physics, **0605** (2006), 026.

[12] T. Sjostrand, et al. *A Brief Introduction to PYTHIA 8.1.* Computer Physics Communications, **178** (2008), 852–867.

[13] PDG. *Review of particle physics.* Physical Review, **D86** (2012) 1, 010001.

[14] LHCb Collaboration. *The LHCb Detector at the LHC.* Journal of Instrumentation, **3** (2008) 08, S08005.

[15] P. R. Barbosa-Marinho et al. *LHCb VELO (VErtex LOcator): Technical Design Report*. CERN-LHCC-2001-0011 (2001).

[16] S. Amato et al. *LHCb magnet: Technical Design Report*. CERN-LHCC-2000-007 (2000).

[17] P. R. Barbosa-Marinho et al. *LHCb inner tracker: Technical Design Report*. CERN-LHCC-2002-029 (2002).

[18] P. R. Barbosa-Marinho et al. *LHCb outer tracker: Technical Design Report*. CERN-LHCC-2001-024 (2001).

[19] M. Needham. *Performance of the LHCb Track Reconstruction Software*. CERN-LHCb-2007-144 (2008).

[20] L. Collaboration. *Tracking strategies used in LHCb*. Accessed May 2013.

[21] E. Rodrigues. *Dealing with clones in the tracking*. CERN-LHCb-2006-057 (2006).

[22] M. Needham. *Identification of Ghost Tracks using a Likelihood Method*. CERN-LHCb-2008-026 (2008).

[23] G. Lanfranchi, et al. *The Muon Identification Procedure of the LHCb Experiment for the First Data*. (2009) LHCb-PUB-2009-013. CERN-LHCb-PUB-2009-013.

[24] LHCb Collaboration. *Absolute luminosity measurements with the LHCb detector at the LHC*. Journal of Instrumentation, **7** (2012), P01010.

[25] M. Williams, et al. *The HLT2 Topological Lines*. CERN-LHCb-PUB-2011-002 (2011).

[26] V. V. Gligorov, et al. *The HLT inclusive B triggers*. LHCb-INT-2011-030 (2011).

[27] R. Aaij, et al. *The LHCb Trigger and its Performance in 2011*. Journal of Instrumentation, **8** (2013), P04022.

Chapter 3

The First Evidence of $B_s^0 \to \mu^+\mu^-$

The first search for $B^0 \to \mu^+\mu^-$ appears in the literature in 1984 [1]. Since then, this channel, together with $B_s^0 \to \mu^+\mu^-$, has been extensively studied, and the experimental upper limits have constantly improved as the main limitation was only statistics. Figure 3.1 and Figure 3.2 show this progress and pay tribute to CDF and CLEO which led for more than a decade each the searches for the $B_s^0 \to \mu^+\mu^-$ and the $B^0 \to \mu^+\mu^-$ decays respectively. These Figures show as well that the constant exponential improvement of the upper limits achieved by these two experiments is nowadays taken over by CMS and LHCb with exponential slopes five time steeper. This impressive acceleration is mainly due to the very high luminosity and production cross section achieved at the LHC (see Section 2.1).

This chapter describes the LHCb analysis which led to the first observation of the $B_s^0 \to \mu^+\mu^-$ decay. The chapter starts with Section 3.1 by an overview of the analysis in order to settle the key elements used in the following. This section will explain that the observed events are classified in categories according to their probability to be signal or background. The sensitivity of the analysis depends strongly on the choice of these categories whose optimisation is described in Section 3.2. Then Section 3.3 describes the software developed to reproduce the analysis process on a large amount of pseudo-experiments. This software is essential as it allows to make statements with confidence levels on the value of the branching fractions. The results obtained with this software are discussed in Section 3.4 (however theoretical interpretations will only be derived in Section 4). The previous results, in particular those brought by the CMS Collaboration, are still relevant and large improvements can be achieved by combining them with the new results. Section 3.3 describes this combination and its results.

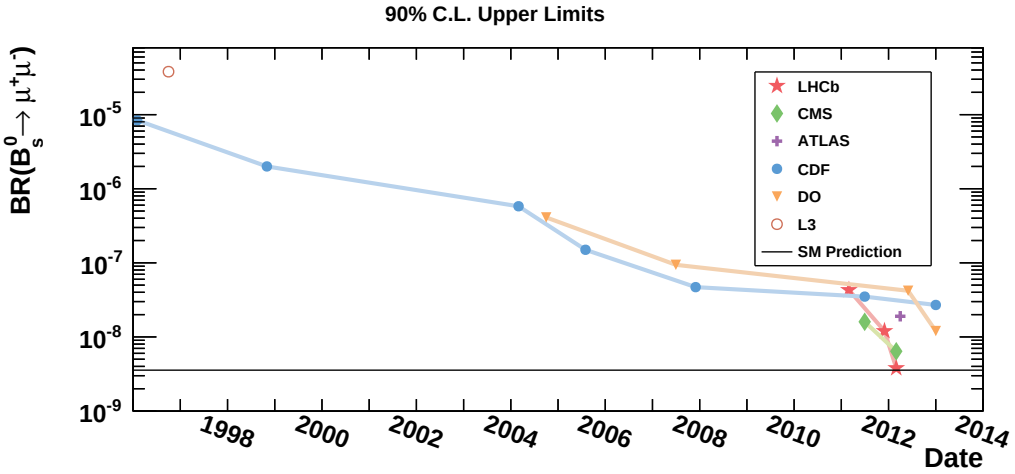


Figure 3.1: Upper limits at 90% C.L. on $\mathcal{B}(\mathrm{B}_\mathrm{s}^0 \to \mu^+\mu^-)$ since the first analysis in 1996. The results of the analysis described in this thesis are not included.

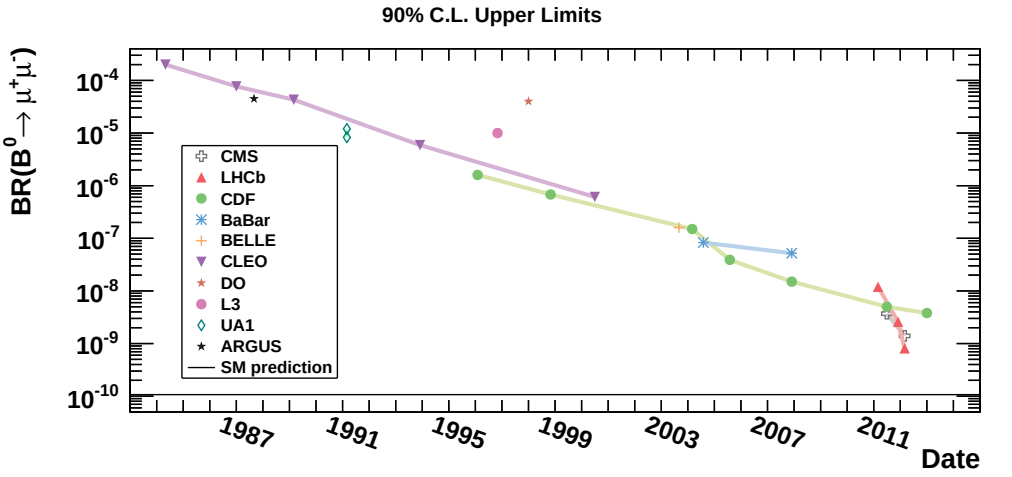


Figure 3.2: Upper limits at 90% C.L. on $\mathcal{B}(\mathrm{B}^0 \to \mu^+\mu^-)$ since the first analysis in 1984. The results of the analysis described in this thesis are not included.

3.1 Overview of the Analysis

In this section the main steps of the analysis are described in the order in which they are performed in practice, as this order is the most natural one. However, some analysis choices in early steps are motivated by considerations appearing in later stages. Hence, in order to describe coherently these choices, the following introduction sketches the broad lines of the analysis.

3.1.1 Introduction

The branching fraction of $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ is defined as the ratio of the number of B_s^0 decaying to two muons by the total number of B_s^0 produced. The aim of the analysis is therefore to extract these two numbers from the data sample selected by various triggers.

The trigger selection is designed to be very efficient for $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ signal. The counter part of this high signal efficiency is the contamination of the sample by backgrounds events. These fake signal events stem from five main sources:

combinatorial prompt candidates, for which at least one of the two muons forming the candidate originates from the primary vertex,

combinatorial long-lived candidates, for which the μ^- and the μ^+ originate from semi-leptonic decays of the $\overline{\mathsf{b}}$ and b which are always produced in pairs, as illustrated in Figure 3.3,

B **cascade candidates,** ($\mathsf{B} \to \mathsf{D}(\to \mu\mathsf{X})\mu\mathsf{X}$), where one of the muon originates from the decay of a B to a charmed meson that subsequently decays and produces the other muon to form the signal candidate,

misidentified $\mathsf{B}^0 \to \mathsf{h}^+\mathsf{h}'^-$ **candidates,** for which the two hadrons decay in flight each to a muon and a neutrino and are therefore identified as muons,

other exclusive decays candidates, which mainly come from $\mathsf{B}^{0,+} \to \pi^{0,+}\mu^+\mu^-$ and $\mathsf{B}^0 \to \pi^-\mu^+\bar{\nu}_\mu$ with a misidentified pion, or from the di-muons pairs produced by the $\mathsf{p}^+\mathsf{p}^+ \to \mathsf{p}^+\mu^-\mu^-\mathsf{p}^+$ process.

The backgrounds levels determine the analysis sensitivity as the magnitude of the statistical fluctuations of the expected number of background events fixes the level at which an excess with respect to the expected number of background events can be attributed to a $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ signal. Quantitatively the sensitivity is usually measured by the number of signal events in

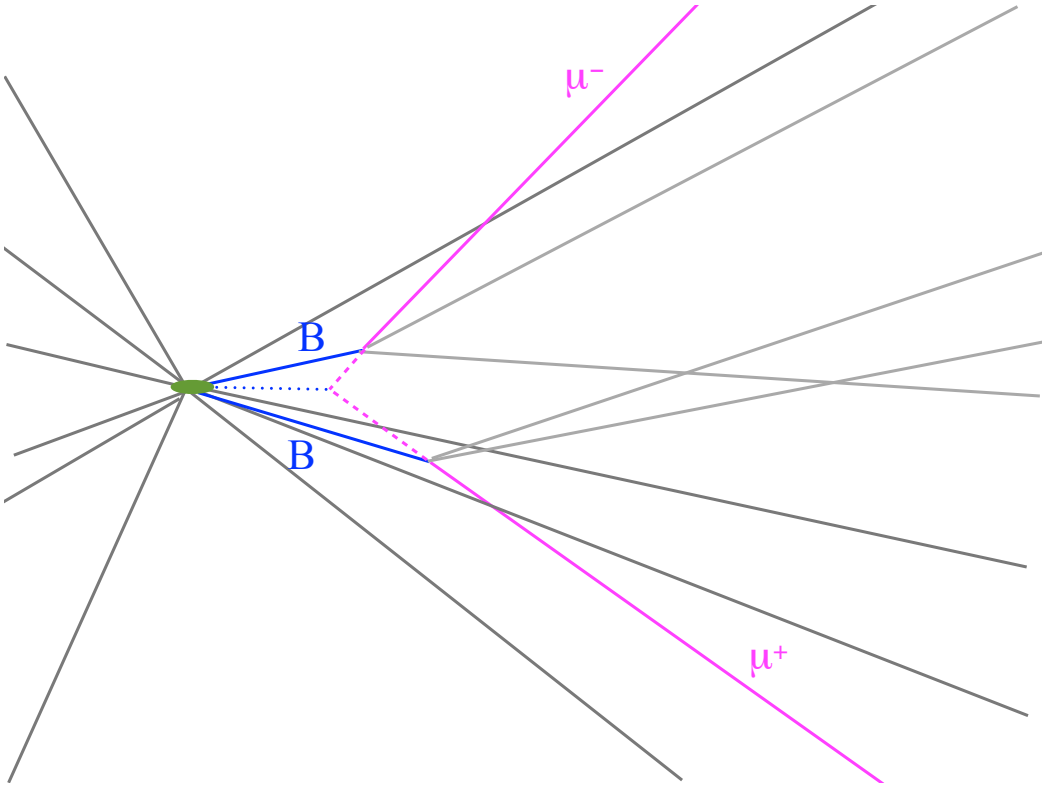


Figure 3.3: Cartoon of a typical combinatorial candidate. The green ellipse is the primary vertex. The two pink lines are muons which originate from the decays of two B mesons (blue). Their extrapolations (dashed pink) cross and make the B candidate vertex (dashed blue).

units of standard deviation of the number of expected signal and background events, $N_S/\sqrt{N_B + N_S}$, usually referred as significance. Hence, it is of prime interest to reduce the background keeping as much signal as possible. For this purpose, a first loose selection is applied which reduces the background by 70% for a signal efficiency of 93%.

At this stage, any additional selection would remove a part of the sample in which the signal events have already a non negligible significance. Therefore, to further increase the sensitivity, we choose to classify events in categories of increasing signal purity. These categories are defined with the di-muon invariant mass and a multivariate operator combining information on the topology and kinematics of the candidate. The sensitivity gain achieved by this classification process depends on the discrimination power of the multivariate classifier and on the category designed which is described in Section 3.2.

The methods to determine the number of background and signal events in each category are designed to rely preferably on the data instead of simulation. The relative bin signal contents are obtained using control channels ($\mathsf{B}^0 \to \mathsf{h}^+\mathsf{h}'^-$ and $\mathsf{B}^0_\mathsf{s} \to \mathsf{h}^+\mathsf{h}'^-$) which have a similar topology. The selection and classification described previously must therefore be similar between signal and control channels.

The second number to be derived is the total number of B^0_s produced. This quantity is obtained indirectly by measuring the number of events in channels (called later normalisation channels, $\mathsf{B}^+ \to \mathsf{J}/\psi\mathsf{K}^+$ and $\mathsf{B}^0 \to \mathsf{K}^+\pi^-$), for which the branching fraction has been measured by other experiments, and correcting for the probability that a b quark hadronises with an s quark instead of a d quark. Once again, the selection performed in the early stage of the analysis must be designed to behave similarly on signal and on normalisation channels.

3.1.2 Simulated Data Sample

The analysis relies on simulated data in particular to define the classifier used to separate signal and background and to understand specific backgrounds. The simulation of these data is performed in three main steps, the event generation, the detector simulation, and the event reconstruction.

The event generation is handled with a specific software called Gauss [2]. This software relies on Pythia [3, 4] to simulate the hard processes happening at the particle collision and the subsequent hadronisation. At the end of this first stage, particles are produced with a given energy and direction. The decay of these particles is simulated with EvtGen [5], which relies itself on

other software like Photos to simulate the final state radiation[1]. Then the particles are propagated and their interaction with the detector simulated with Geant4 [6]. A substantial amount of CPU time can be saved by requiring the particles of interest to have a momentum direction within the LHCb acceptance.

The detector response to this particle flow is simulated in a second step with Boole [7]. In addition Boole simulates spill-over effects[2] and LHC backgrounds. Its output is digitised to mimic real data coming from the detector.

Finally, Brunel [8] performs the event reconstruction and the trigger is emulated with Moore [9]. The genealogy of the reconstructed particle can be recovered, hence the true origin of the particles can be identified.

The list of processes simulated for the analysis is shown in Table 3.1. The $\mathsf{b\bar{b}} \to \mu\mu\mathsf{X}$ samples contain events in which two B mesons are produced and decay into final states with at least one muon. This type of events is at the origin of the combinatorial background candidates illustrated in Figure 3.3.

3.1.3 Trigger

The data used in the analysis were recorded with five TCKs corresponding in fact to three different sets of selection criteria, TCK1, TCK2, and TCK3. $\mathsf{B}^0_{(\mathsf{s})} \to \mu^+\mu^-$ and $\mathsf{B}^+ \to \mathsf{J}/\psi\mathsf{K}^+$ events are selected by similar trigger lines, as their most distinctive features is the muons pair. $\mathsf{B}^0_{(\mathsf{s})} \to \mathsf{h}^+\mathsf{h}'^-$ events are selected by different lines but similar selection criteria.

At L0 $\mathsf{B}^0_{(\mathsf{s})} \to \mu^+\mu^-$ and $\mathsf{B}^+ \to \mathsf{J}/\psi\mathsf{K}^+$ events are selected mainly by `L0SingleMuon` and `L0Dimuons` while $\mathsf{B}^0_{(\mathsf{s})} \to \mathsf{h}^+\mathsf{h}'^-$ events are selected by `L0Hadron`. These trigger lines have very similar selection criteria which are reported for each TCK in Table 2.2.

At HLT1 $\mathsf{B}^0_{(\mathsf{s})} \to \mu^+\mu^-$ and $\mathsf{B}^+ \to \mathsf{J}/\psi\mathsf{K}^+$ events are selected by the three muonic lines, `HLT1TrackMuon`, `Hlt1DiMuonLowMass` and `Hlt1DiMuonHighMass` and $\mathsf{B}^0_{(\mathsf{s})} \to \mathsf{h}^+\mathsf{h}'^-$ events by `Hlt1TrackAllL0`. Like for the L0, these trigger lines are very similar and their selection criteria as reported for each TCK in Table 2.3.

[1]The simulation of the final state radiation is a crucial issue to have a reliable comparison between experimental results and theoretical predictions. The signal simulated data were generated with Photos v215.4.

[2]Spill-over effects are detector effects which originate from physical processes lasting longer than the time span between two bunch crossing (e.g. hits originating from slow particles or particle drift in the outer tracker).

Table 3.1: Monte Carlo samples used in this analysis. Samples generated with MC2012 (MC10 or MC11a) assume a pp interation at 8 (7) TeV.

Channel	Monte Carlo Production	Magnet Polarity	Events Processed	Events Rec. and Stripped
Signal:				
$\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$	MC2012sim06	Down / Up	127 k	45k
Combinatorial background:				
$\mathsf{b\bar{b}} \to \mu\mu\mathsf{X}$	MC10sim01	Down	50 M	57 k
$\mathsf{b\bar{b}} \to \mu\mu\mathsf{X}$	MC10sim01	Up	50 M	57 k
Control/normalization channels:				
$\mathsf{B}_\mathsf{s}^0 \to \mathsf{J}/\psi\phi$	MC2012	Down/Up	127k	14k
$\mathsf{B}^+ \to \mathsf{J}/\psi\mathsf{K}^+$	MC2012	Down/Up	127k	25k
$\mathsf{B}^0 \to \mathsf{K}^+\pi^-$	MC2012	Down/Up	207k	
Exclusive backgrounds				
$\mathsf{B}^0 \to \pi^-\mu^+\nu_\mu$	MC11a	Down/Up	5 M	–
$\mathsf{B}_\mathsf{s}^0 \to \mathsf{K}^-\mu^+\nu_\mu$	MC11a	Down/Up	5 M	–
$\mathsf{B}^+ \to \pi^+\mu^+\mu^-$	MC11a	Down/Up		–
$\mathsf{B}^+ \to K^+\mu^+\mu^-$	MC11a	Down/Up	1 M	–
$\Lambda_\mathsf{b} \to p\mu^-\bar{\nu}_\mu$	MC11a	Down/Up	5 M	–
$\mathsf{B}_\mathsf{c}^+ \to J/\psi(\to \mu^+\mu^-)\mu^+\nu_\mu$	MC11a	Down/Up	5 M	–

At HLT2 $B^+ \to J/\psi K^+$ events are selected in the first $630\,\text{pb}^{-1}$ by `Hlt2DiMuonJPsi` then in the last $470\,\text{pb}^{-1}$ this line was pre-scaled and therefore these events were mostly selected by `Hlt2DiMuonDetached`. $B_s^0 \to \mu^+\mu^-$ events are mainly selected by `Hlt2DiMuonBmm`. And $B^0_{(s)} \to h^+h'^-$ events are mainly selected by `Hlt2Topo2Body`, `Hlt2B2hhX`, and `Hlt2B2hh` The selection criteria of these lines are very similar and reported in Table 2.5.

3.1.4 Initial Selection

The initial purpose of the selection is to gather good signal candidates by applying some quality requirements and remove obvious backgrounds. The core of the discrimination between signal and background is left to the multivariate classifier described in Section 3.1.6. The selection design must behave similarly on all channels used for the analysis: $B^0_{(s)} \to \mu^+\mu^-$, $B^+ \to J/\psi K^+$, and $B^0_{(s)} \to h^+h'^-$. Therefore in the following paragraphs *signal* is used to refer to any of these channels. The criteria chosen to achieve this purpose and match this constraint are described in the following paragraphs and the numerical values of the selection criteria are reported in Table 3.2.

To pass the selection, candidates must be made with tracks with a good χ^2, a small probability to be ghost tracks (i.e. tracks made of hits not deposed by a single particle), and a small Kullback-Leibler (KL) distance which allows to remove tracks duplicated at the reconstruction level [10, 11, 12]. In addition to these single track quality criteria, the tracks must have a small distance of closest approach (DOCA) between each other and the two hadrons tracks for $B^0_{(s)} \to h^+h'^-$ and the two muon ones for $B^0_{(s)} \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ must form a vertex with a good χ^2.

Aside from these quality requirements, obvious backgrounds are removed by additional selection criteria. The B mesons produced from the 7 TeV pp collisions are highly boosted and fly, in average, 1 cm away from the primary vertex. This feature allows to reject most of the *prompt combinatorial backgrounds* by requiring the distance significance between the B candidate decay point or the J/ψ one for $B^+ \to J/\psi K^+$ and the primary vertex (VDS) to be large enough and requesting the B candidate to have a small impact parameter significance (IPχ^2) with respect to the primary vertex but to be made of tracks with a large one. The selection takes also advantage of the signal kinematics to further reduce the backgrounds by asking the candidate to have a invariant mass within 500 MeVof the central expected value of the B meson mass and within 60 (100) MeV/c^2 of the J/ψ and (B^+) one for $B^+ \to J/\psi K^+$.

The backgrounds in the muonic channels due to candidates made of

misidentified hadrons are reduced with the particle identification variables IsMuon, described in [13], and $\mathsf{DLL}(\mu,\pi)$, $\mathsf{DLL}(\mathsf{K},\pi)$. The two last variables are the logarithms of the ratio between the likelihood for the particle to be a μ (or K) and a π.

The *exclusive di-muons processes* $\mathsf{p}^+\mathsf{p}^+ \to \mathsf{p}^+\mu^-\mu^-\mathsf{p}^+$ can produce candidates very similar to the signal when the two protons fall in the beam pipe and if the candidate made of the muon pair points to a primary vertex. However the candidates produced by these processes present a very soft transverse momentum spectra [14] and are therefore easily removed by requiring the candidates to have a minimum transverse momentum of $0.5\,\mathrm{GeV}/c$ with respect to the beam pipe.

Background candidates could also originate from $\mathsf{B}^0_{(\mathsf{s})} \to \ell^+\ell^-\gamma$ processes, i.e. processes where the photon is not emitted in the final state. This pollution can be substantial as the additional photon removes the helicity suppression, making the branching fractions of these processes larger than $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ ones. However, the photons have such large momenta that the resulting B candidates do not pass the invariant mass selection and, its contribution is negligible. For example, for $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ the effective branching fraction of these processes in the signal's invariant mass region ($\pm 60\,\mathrm{MeV}$) is estimated to be three orders of magnitude smaller than $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ [15]:

$$\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-\gamma)_{E(\gamma)<60\,\mathrm{MeV}} \;=\; 1.6 \times 10^{-12}. \tag{3.1}$$

Finally the analysis is protected against *non-physical* candidates by removing all candidates which have either a momentum, a transverse momentum, or a lifetime too large. This fiducial selection is designed to be 99.9% efficient for the signal.

3.1.5 Tight Selection and BDT

After the initial selection, the prompt, misidentified, and exclusive backgrounds have been reduced, however the sample still contains a large amount of long-lived combinatorial background. The small signal over background ratio is depicted for the $\mathsf{B}^0_{(\mathsf{s})} \to \mathsf{h}^+\mathsf{h}'^-$ control channel in Figure 3.4, which shows the mass spectrum of candidates after the initial selection for the whole sample and for the sub-sample of the events unbiased by the trigger selection. The signal yield of this sub-sample needs to be extracted to derive the relative classifier bin contents (see Section 3.1.7), and such an extraction is possible only if the signal purity is increased. In order to keep as much signal events as possible, we require any tighter selection to have a signal efficiency of at least 95%. Matching this criterion by tightening the selection

Table 3.2: Selection for $\mathsf{B}^0_{(s)} \to \mu^+\mu^-$, $\mathsf{B}^0_{(s)} \to \mathsf{h}^+\mathsf{h}'^-$, and $\mathsf{B}^+ \to \mathsf{J}/\psi\mathsf{K}^+$ channels.

cut	applied on	value for $\mathsf{B}^0_s \to \mu^+\mu^-$, $\mathsf{B}^0_{(s)} \to \mathsf{h}^+\mathsf{h}'^-$	applied on	value for $\mathsf{B}^+ \to \mathsf{J}/\psi\mathsf{K}^+$
track χ^2/ndf	μ / h	< 3	μ / h	< 3
ghost prob		< 0.4373		< 0.4373
DOCA		$<0.3\,\mathrm{mm}$		$<0.3\,\mathrm{mm}$
IPχ^2		>25		>25
p_T		> 0.25 GeV/c		> 0.25 GeV/c
p_T		< 40 GeV/c		< 40 GeV/c
p		$<500\,\mathrm{GeV}/c$		$<100\,\mathrm{GeV}/c$
KL		< 5000		< 5000
IsMuon	μ only	true	μ only	true
vertex χ^2	$B_{(s)}$	<9	J/ψ	<9
VDS		> 15		> 15
ΔM		<500 MeV/c^2		$< 500\,\mathrm{MeV}/c^2$
IPχ^2	$B_{(s)}$	< 25	B^+	< 25
t		$< 9 \cdot\tau(\mathsf{B}^0_s)$		$< 9 \cdot\tau(\mathsf{B}^0_s)$
BDTS		> 0.05		> 0.05
DLL(K, π)		< 10		
DLL(μ, π)		> -5		
ΔM				$< 500\,\mathrm{MeV}/c^2$
$p_T(\mathsf{B}^0_s)$	$\mathsf{B}^0_s \to \mu^+\mu^-$	$> 0.5\,\mathrm{GeV}/c$		

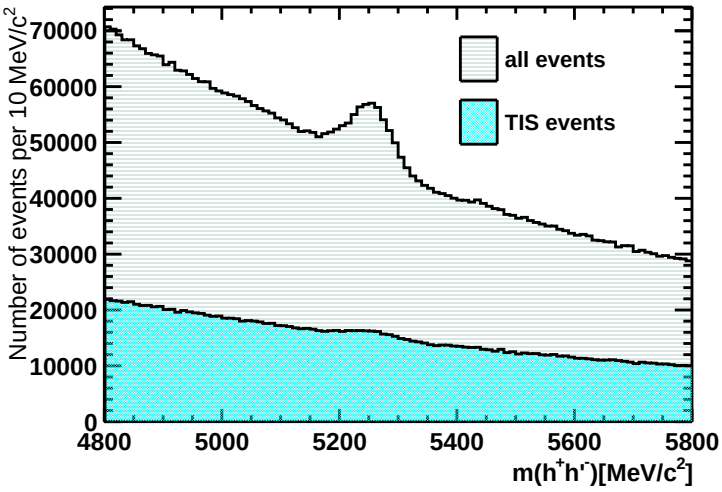


Figure 3.4: Invariant mass distribution of a $B^0_{(s)} \to h^+h'^-$ $800\,pb^{-1}$ sample (dashed area) and of its trigger unbiased (TIS) events (blue area).

criteria described in Section 3.1.4, yields a background rejection of 50% while a Boosted Decision Tree (BDT) multivariate selection, called BDTS, allows to reach a 70% background rejection.

Boosted Decison Tree

Boosted Decision Trees (BDT) are techniques which are designed to classify events, described with a list of variables, in several categories, signal and background in our case. Their designs rely on two samples of pure background and pure signal events from which the BDT learn to discriminate between each of the two species. With these two samples the algorithm looks for the single variable based selection criterium which maximises the separation between signal and background. The samples are then split into two sub-samples, called leaves, and the process is repeated in each sub-sample. After n iterations, this process has built a decision tree (DT) of depth n and the n^{th} leaves of the tree are tagged signal or background according to the specie they mainly contain. Such a tree is shown in Figure 3.5. An event passed through the DT will be assigned an output value of zero or one according to the type of the final leaf it lands on. A new tree is build from the same initial mixed sample in which events are weighted in a way that the ones previously misclassified will be more important. This events weighting is usually called *boosting*. Several DT are built in this fashion and their outputs

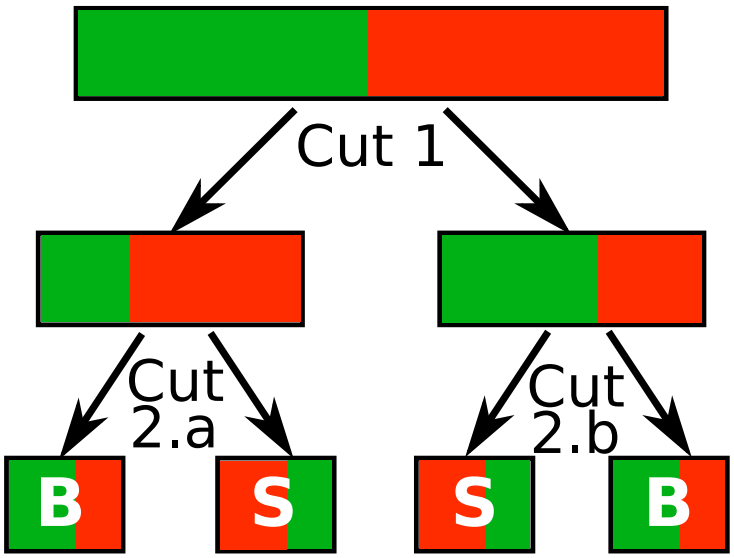


Figure 3.5: Decision tree of depth 2. The green (red) symbolises the background (signal) component of the mixed data set. The four last leaves are labelled S for signal and B for background.

are finally combined by taking the weighted average of the single DT output. Finally we reshape the BDT output to have a flat distribution between 0 and 1 for the signal, resulting into a peaked distribution for the background at low BDT values. This reshaping allows to convert any signal efficiency in a selection criteria. At the end of the building process the BDT is tested on another set of signal and background samples statistically independent from the ones used in the training. Any loss in the BDT performances on these samples indicates that the classifications is operated on event features that are no longer statistically relevant.

The Boosted Decison Tree based selection

The BDT used to perform the tighter selection is built from the six following variables:

- the impact parameter[3] and impact parameter χ^2 of the B candidate,
- the χ^2 of the B meson or of the $\mathsf{J}\!/\!\psi$ vertex for $\mathsf{B}^+ \to \mathsf{J}\!/\!\psi\mathsf{K}^+$,
- the angle between the direction of the momentum of the B candidate and the direction of the vector linking the primary to the B vertex,

[3]common with the BDT used later for the event classification, see Section 3.1.6.

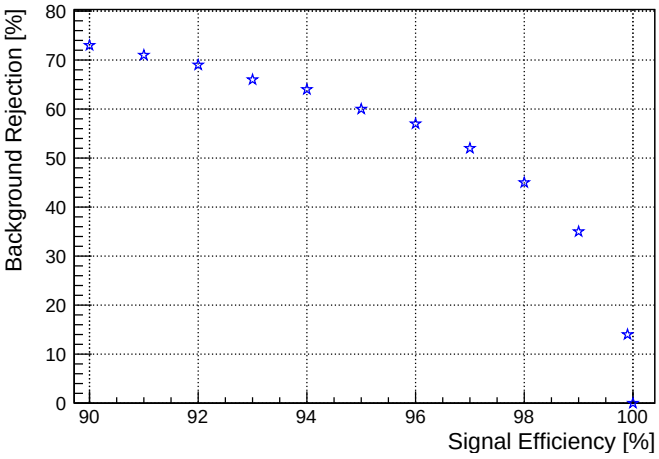


Figure 3.6: Background rejection versus signal efficiency for a BDT based selection computed with simulated $B_s^0 \to \mu^+\mu^-$ as signal and $B^0_{(s)} \to h^+h'^-$ data sidebands as background.

– the minimum distance between the daughter tracks (only the two muons for $B^+ \to J/\psi K^+$) [3],

– the minimum impact parameter of the daughter tracks with respect any primary vertex.

The BDT is trained on simulated data for the signal and on the signal sidebands $[4800 - 5000] \cup [5500 - 6000]\,\mathrm{MeV}/c^2$ of the di-hadrons mass spectra for the background[4]. The BDT performances are summarised in the curve of background rejection versus signal efficiency in Figure 3.6. This curve shows that for a 93% signal efficiency the BDT based selection rejects 70% of the background, increasing clearly the signal over background ratio of the trigger unbiased $B^0_{(s)} \to h^+h'^-$ sample, as illustrated in Figure 3.7.

The scatter plot of the mass versus the BDT shown in Figure 3.8 demonstrates that the two variables are not correlated.

Finally, the efficiencies are found to be consistent between signal, normalisation, and control channels within 0.4%.

[4]This region of the $B^0_{(s)} \to h^+h'^-$ spectrum is not used afterwards in the analysis so training the BDT on these events implies no bias.

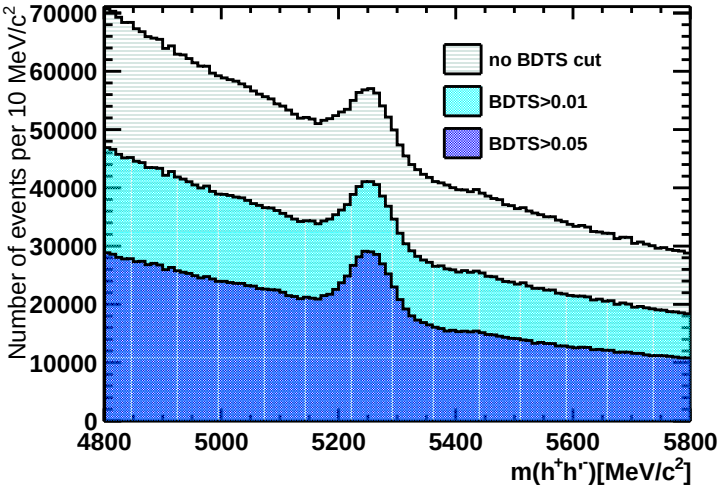


Figure 3.7: $B^0_{(s)} \to h^+h'^-$ invariant mass distributions before (dashed area) and after the selection based on the BDTS (dark blue area).

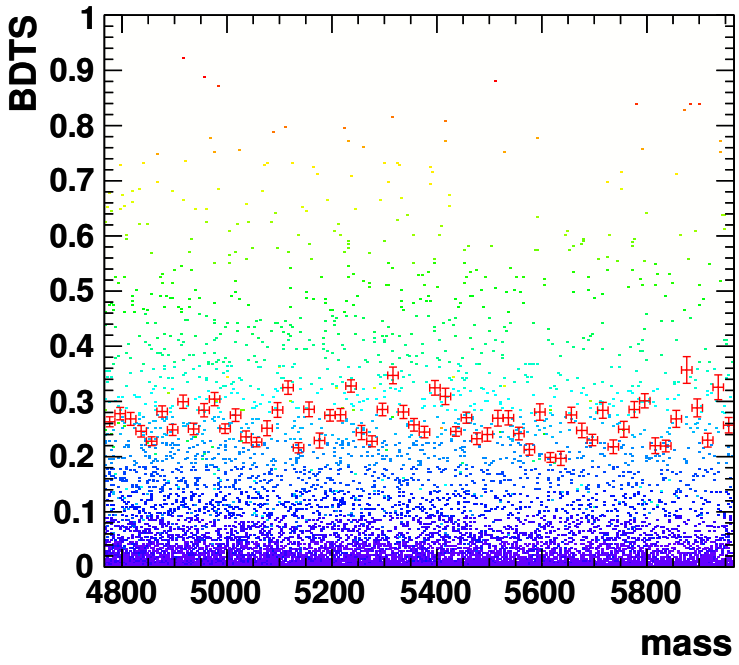


Figure 3.8: BDT versus the invariant mass for simulated $b\bar{b} \to \mu\mu X$. The red cross are the BDT averages for each invariant mass bin.

3.1.6 Classification

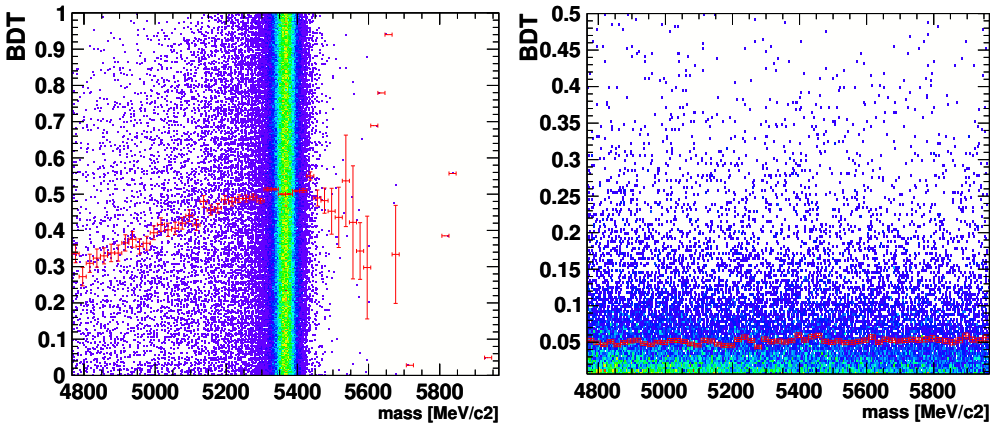
At this level of the analysis, the signal purity of the sample has been increased to a level at which, in any sub-sample selected by any of the best available multivariate classifiers, the SM signal events would have a non negligible significance with respect to the background. Hence, in order not to loose any sensitivity, no further selection is applied. Instead, events are classified in bins of the candidate invariant mass and of a topological based multivariate discriminator and these bins are combined. The optimisation of the bin boundaries is described in details in Section 3.2.

The discriminator used for this classification is also a BDT. This BDT is trained and tested on two independent samples of simulated $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ events for the signal and on simulated $\mathsf{b\bar{b}} \to \mu\mu\mathsf{X}$ events for the background. The optimisation of the set of input variables [16] selects the nine following variables:

- B meson proper time (t),
- minimum impact parameter significance of the muons ($IPS(\mu)$) with respect to any primary vertex,
- the impact parameter of the B ($IP(B)$) with respect to the primary vertex,
- distance of closest approach between the two muons ($DOCA$),
- the sum of the isolation of the two muons with respect to any other track in the event ($I(\mu)$, the number of good two-track vertices a muon can make with other tracks in the event),
- the transverse momentum of the B meson ($p_T(B)$),
- the cosine of the angle between the muon momentum in the dimuon rest frame and the vector perpendicular to the B momentum and the beam axis ($\cos P$),
- the B isolation based on the CDF definition ($I(B)$) [17]:

$$I(B) \;=\; \frac{p_T(B)}{p_T(B) + \sum_{tracks} p_T(tracks)},$$

- the minimum transverse momentum of the two muons ($p_{T,\min}(\mu)$).



(a) (b)

Figure 3.9: Scatter plots of the invariant mass versus the BDT output for $B_s^0 \to \mu^+\mu^-$ MC event (a) and for $b\bar{b} \to \mu\mu X$ (b). The red crosses are the BDT averages for each invariant mass bin.

The BDT must be designed to have an output uncorrelated (at worse linearly correlated) with the invariant mass of the candidate otherwise the method used to extrapolate the background content of the signal mass region content from the sidebands, described in Section 3.1.7, would be biased. Figure 3.9 shows no correlation between the invariant mass and the BDT for the combinatorial background while it exhibits a 6% correlation for the signal over the whole mass range. This correlation is due to the radiative decays at the low edge of the mass distribution.

3.1.7 Classifier Calibration

Once the bi-dimensional binned classification scheme has been established, the signal and background contents for each classification category must be derived. This is done as much as possible with data driven method, to avoid relying on simulated data.

The signals invariant mass distributions are assumed to be Crystal Ball functions [18] to account for their radiative tails. The means of the Gaussian cores and their resolutions are obtained from data. The former are derived by fitting the $B^0_{(s)} \to h^+h'^-$ invariant mass distributions. These fits provide as well estimates for the invariant mass resolutions which are combined with the ones obtained by interpolating, with a power law, at the B_s^0 and B^0 mass, the resolutions measured at the quarkonium resonances (J/ψ, $\psi(2S)$, $\Upsilon(1,2,3S)$). The mean and resolution estimates for the 2012 analysis [19]

are shown in Table 3.3 which clearly demonstrates that the two modes are resolved. Finally the position of the transition point between the Gaussian core and the exponential tail, is estimated with simulated data and is found to be at $(2.094 \pm 0.028)\sigma_{m_{\mu\mu}}$ of the Gaussian mean.

Table 3.3: Invariant mass means and resolutions for the two signal modes.

Mode	Mean, MeV/c^2	Resolution, MeV/c^2
B^0_s	$(5371.55 \pm 0.41_{\text{stat}} \pm 0.16_{\text{syst}})$	$(25.04 \pm 0.18_{\text{stat}} \pm 0.36_{\text{syst}})$
B^0	$(5284.36 \pm 0.26_{\text{stat}} \pm 0.13_{\text{syst}})$	$(24.63 \pm 0.13_{\text{stat}} \pm 0.36_{\text{syst}})$

The signal BDT output is designed to be flat between 0 and 1, so that we can translate directly any selection criteria in term of signal efficiency. The background is therefore peaked at low BDT values. For the signal, the BDT PDF is obtained using $\mathsf{B}^0_{(\mathsf{s})} \to \mathsf{h}^+\mathsf{h}'^-$ decays as they have the same topology as $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$. To minimise the bias due to different requirements between the trigger lines selecting $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ candidates and those selecting $\mathsf{B}^0_{(\mathsf{s})} \to \mathsf{h}^+\mathsf{h}'^-$, only the events for which the triggers have fired independently of the $\mathsf{B}^0_{(\mathsf{s})} \to \mathsf{h}^+\mathsf{h}'^-$ decay products are used. Then in each BDT bin, the $\mathsf{B}^0_{(\mathsf{s})} \to \mathsf{h}^+\mathsf{h}'^-$ yields are extracted with a fit to the invariant mass distribution. The invariant mass line shape is described by a sum of two Crystal Ball functions [18], to account for $\mathsf{B}^0_\mathsf{s} \to \mathsf{h}^+\mathsf{h}'^-$ and $\mathsf{B}^0 \to \mathsf{h}^+\mathsf{h}'^-$ an exponential function which describes $\Lambda_\mathsf{b} \to \mathsf{p}\mathsf{h}^-$, where the p is identified as a K^+ or π^+, and a specific PDF describing the partially reconstructed decay [20]:

$$\begin{aligned} f_P(m) &= A^{-1} \times m' \left(1 - \frac{m'^2}{m_0^2}\right) \Theta(m_0 - m') e^{-C_P m'} \\ &\otimes G(m - m'; \sigma_P), \end{aligned} \tag{3.2}$$

where Θ is the step function, $\otimes$ the convolution product (on m'), $G(m - m'; \sigma_P)$ a Gaussian with standard deviation σ_P representing the experimental resolution, m_0 a free parameter, A a normalization factor, and C_P a parameter governing the shape of the distribution, with no obvious physical meaning.

The mass hypotheses of the decay products are attributed on the basis of the two particle identification variables, $\mathsf{DLL}(\mathsf{K}, \pi)$ and $\mathsf{DLL}(\pi, \mathsf{K})$, and the events are weighted according to the identification efficiencies of these two variables. An example of this fit is shown in Figure 3.10. The resulting BDT PDF is shown in Figure 3.12.

For the background in the $\mu^+\mu^-$ sample, the line shape in the mass range [4900–5996] MeV/c^2 is the sum of an exponential function for the com-

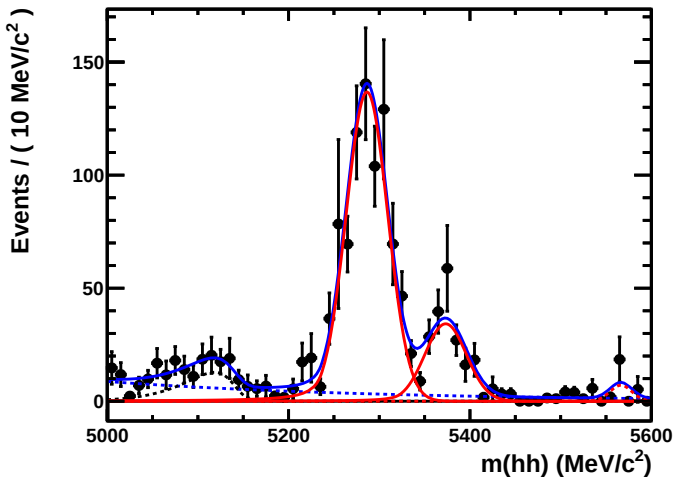


Figure 3.10: Invariant mass distribution of $B^0_{(s)} \to h^+h'^-$ candidates in the last BDT bin [0.9–1]. The red solid lines show the $B^0 \to h^+h'^-$ and $B^0_s \to h^+h'^-$ signals, and the red dashed one the $\Lambda_b \to ph^-$ signal. The combinatorial background is shown by the blue dashed line, while the physical background (partially-reconstructed) is shown by the black dashed line. Hadrons are identified as kaons if $\mathsf{DLL}(K, \pi) \geq 5$, as pions if $-\mathsf{DLL}(K, \pi) \geq 5$ and rejected if $|\mathsf{DLL}(K, \pi) \leq 5|$.

binatorial backgrounds, a double Gaussian function for the misidentified $B^0_{(s)} \to h^+h'^-$[5] and a simple Gaussian function for $B^0 \to \pi^-\mu^+\bar{\nu}_\mu$ and $B^{0,+} \to \pi^{0,+}\mu^+\mu^-$. The expected number of background events in each BDT bin, and therefore the BDT PDF, is obtained by fitting the sidebands and extrapolating the results of the fit into the signal mass windows. An example of this fit is illustrated in Figure 3.11. The resulting BDT PDF is shown in Figure 3.12.

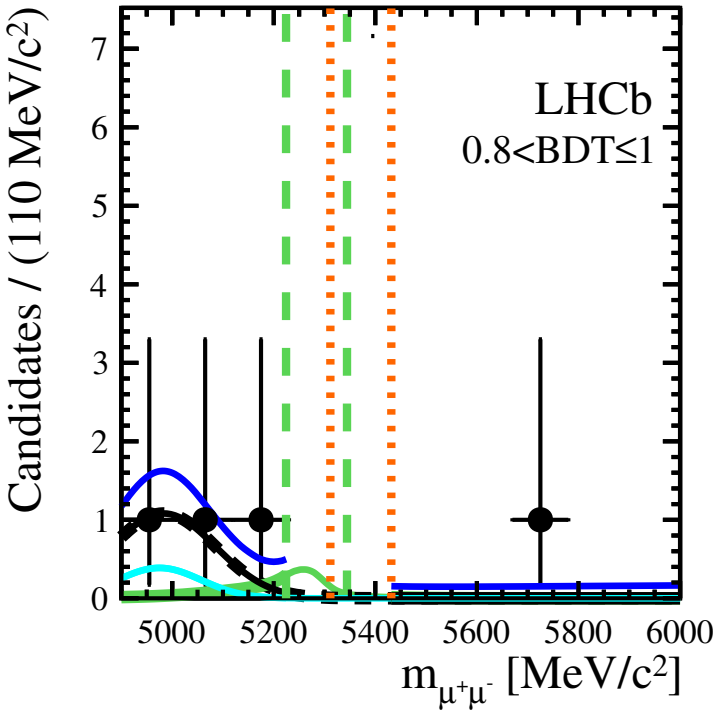


Figure 3.11: Invariant mass distribution of 2012 data sidebands in the last two BDT bins [0.8–1]. Dots are data, the black line is the $B^0 \to \pi^-\mu^+\bar{\nu}_\mu$, the cyan line the $B^{0,+} \to \pi^{0,+}\mu^+\mu^-$, the green line the misidentified $B^0 \to h^+h'^-$, and the blue the total fit. Vertical orange (green) dotted (dashed) lines indicate the $B^0_s \to \mu^+\mu^-$ ($B^0 \to \mu^+\mu^-$) search windows excluded from the background estimation fit.

[5]More details on the $B^0_{(s)} \to h^+h'^-$ mass PDF are given in Section 3.3.2

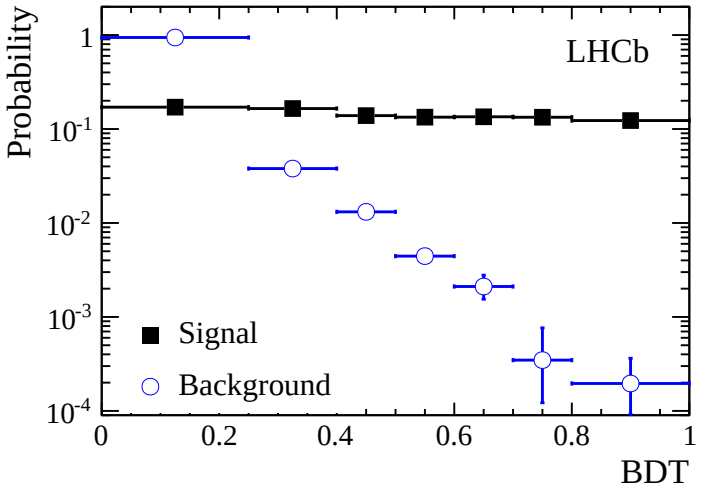


Figure 3.12: Signal (black squares) and background (blue circles) BDT PDF after calibration.

3.1.8 Normalisation

Translating the number of observed signal events, $N(\mathrm{B}^0_\mathrm{s} \to \mu^+\mu^-)$, into a branching fraction requires the total number of B^0_s produced, $N(\mathrm{B}^0_\mathrm{s})$, as $\mathcal{B}(\mathrm{B}^0_\mathrm{s} \to \mu^+\mu^-) \propto N(\mathrm{B}^0_\mathrm{s} \to \mu^+\mu^-)/N(\mathrm{B}^0_\mathrm{s})$. This latter number is obtained by extracting the yields of channels for which the branching fraction has been previously measured and correcting for the efficiency differences between $\mathrm{B}^0_\mathrm{s} \to \mu^+\mu^-$ and such normalisation channels as well as for the probability difference that the b quark hadronises into the signal meson, $\mathrm{B}^0_{(\mathrm{s})}$, or the normalisation channels ones. We use two modes $\mathrm{B}^+ \to \mathrm{J}/\psi\mathrm{K}^+$ and $\mathrm{B}^0 \to \mathrm{K}^+\pi^-$, as their branching fractions are precisely measured, $\mathcal{B}(\mathrm{B}^+ \to \mathrm{J}/\psi\mathrm{K}^+)=(6.01 \pm 0.21) \times 10^{-5}$ and $\mathcal{B}(\mathrm{B}^0 \to \mathrm{K}^+\pi^-)=(1.94 \pm 0.06) \times 10^{-5}$ [21] and they are abundant: in the 2012 data the number of events not biased by the L0 and HLT1 selection are, $N(\mathrm{B}^+ \to \mathrm{J}/\psi\mathrm{K}^+) = 424200 \pm 700$ and $N(\mathrm{B}^0 \to \mathrm{K}^+\pi^-) = 14600 \pm 1100$. In addition these two channels are complementary as the first one has similar trigger and muon identification efficiencies as the signal but is a three body decay, while the second is a two body decay like $\mathrm{B}^0_{(\mathrm{s})} \to \mu^+\mu^-$ but is selected by a different trigger. For each channel, the normalisation factor[6],

[6]The normalisation factor is also called single event sensitivity.

α_{norm} is defined by:

$$\begin{aligned}\mathcal{B}(\mathsf{B}^0_{(s)} \to \mu^+\mu^-) &= \mathcal{B}_{\rm norm} \times \frac{\epsilon^{\rm REC}_{\rm norm}\epsilon^{\rm SEL|REC}_{\rm norm}\epsilon^{\rm TRIG|SEL}_{\rm norm}}{\epsilon^{\rm REC}_{\rm sig}\epsilon^{\rm SEL|REC}_{\rm sig}\epsilon^{\rm TRIG|SEL}_{\rm sig}} \times \frac{f_{\rm norm}}{f_{\mathsf{B}^0_{(s)}}} \times \frac{N_{\mathsf{B}^0_{(s)}\to\mu^+\mu^-}}{N_{\rm norm}} \\ &= \alpha_{\rm norm} \times N_{\mathsf{B}^0_{(s)}\to\mu^+\mu^-}\,, \end{aligned} \tag{3.3}$$

where $f_{\mathsf{B}^0_{(s)}}$ and $f_{\rm norm}$ are the probabilities that a b-quark hadronises into a $\mathsf{B}^0_{(s)}$ and into the b-hadron relevant for the chosen normalisation mode and has been measured in the LHCb kinematics range [22]. This input is the main source of systematic uncertainties and depends on other inputs like for instance $\mathcal{B}(\mathsf{D}^-_s \to \mathsf{K}^+\mathsf{K}^-\pi^-)$. $\mathcal{B}_{\rm norm}$ is the branching fraction and $N_{\rm norm}$ is the number of selected events of the normalisation channel. The efficiency is separated into three factors:

- $\epsilon^{\rm REC}$ is the efficiency to reconstruct all tracks of the decay including the efficiency due to the detector geometrical acceptance,
- $\epsilon^{\rm SEL|REC}$ is the efficiency to select events in which candidates have been reconstructed,
- $\epsilon^{\rm TRIG|SEL}$ is the efficiency to trigger events in which candidates have been reconstructed and selected.

The sub-indexes indicate if the efficiency refers to the signal (sig) or the normalisation channel (norm). Finally, $N_{\mathsf{B}^0_{(s)}\to\mu^+\mu^-}$ is the number of observed signal events.

In Equation 3.3, the efficiencies enter as ratios between normalisation channel and signal efficiencies, therefore all correlated uncertainties between them cancel. Nevertheless, those which are uncorrelated must be precisely evaluated. The following paragraph briefly describes the main methods used to derive these efficiencies.

Reconstruction and Acceptance Efficiency

The reconstruction efficiency results from the detector acceptance and the efficiency to reconstruct all tracks in the final state. For the $\mathsf{B}^0_s \to \mu^+\mu^-$ signal the total reconstruction and acceptance efficiency is 12.7%. Signal and normalisation channel efficiencies are evaluated with simulated data. The majority of the efficiency systematic uncertainties and biases are correlated between signal and normalisation channels and cancel in their ratio. The remaining systematic uncertainties and biases for $\mathsf{B}^+ \to \mathsf{J}/\psi\mathsf{K}^+$ stem from the additional track and the resulting different decay kinematics. For $\mathsf{B}^0 \to \mathsf{K}^+\pi^-$

they arise from the muon system acceptance which is different from the hadron one, from the muon identification, and the particle identification requirements. These uncertainties and biases are all obtained with techniques based on simulation and checked on data.

Selection Efficiency

The selection efficiencies are derived with simulations and the results are checked with data. The selection signal efficiency is 43%. As for the reconstruction efficiency, all uncertainties and biases that are identical between signal and normalisation channels cancel in the ratio. The uncertainties that do not cancel are those related to the ghost probability and the invariant mass selection. These corrections are obtained with a data driven method for the former and with simulation for the latter.

Trigger Efficiency and TISTOS Method

The trigger efficiencies are derived with a data driven method [23]. This method uses the events for which the trigger fired independently of the features of the signal considered. These events are called TIS (Triggered Independently of the Signal). For a given trigger, the efficiency to select any channel can be written as:

$$\begin{aligned} \epsilon &= \frac{N_{trig}}{N_{tot}} \\ &= \frac{N_{trig}}{N_{TIS}} \times \frac{N_{TIS}}{N_{tot}} \\ &= \frac{N_{trig}}{N_{TIS}} \times \epsilon_{TIS}. \end{aligned} \tag{3.4}$$

The probability for an event to be TIS ϵ_{TIS} slightly depends on the channel. Indeed, for TIS events the trigger decision is based on the underlying event features which, in the case of $\mathsf{b\bar{b}}$ events, are correlated to the signal. However, the selection being similar in all channels used for the analysis, this correlation will also be similar and cancel when efficiency ratios are computed.

ϵ_{TIS} , is measured with TIS&TOS events, i.e. events for which both signal and non-signal event features would be sufficient to fire the trigger. As long as the correlations between the signal features and the underlying event can be neglected, the TIS efficiency is given by:

$$\epsilon_{TIS} = \frac{N_{TIS\&TOS}}{N_{TOS}}. \tag{3.5}$$

The $B_s^0 \to \mu^+\mu^-$ trigger efficiency is derived by re-weighting the muon line trigger efficiencies, obtained with the TISTOS method on $J/\psi \to \mu^+\mu^-$, to the simulated $B_s^0 \to \mu^+\mu^-$ maximum transverse momentum and maximum impact parameter distribution. The efficiency is estimated at 90%. Most systematic uncertainties cancel in the efficiency ratio between $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$. Remaining uncertainties come from events triggered by non-muonic lines.

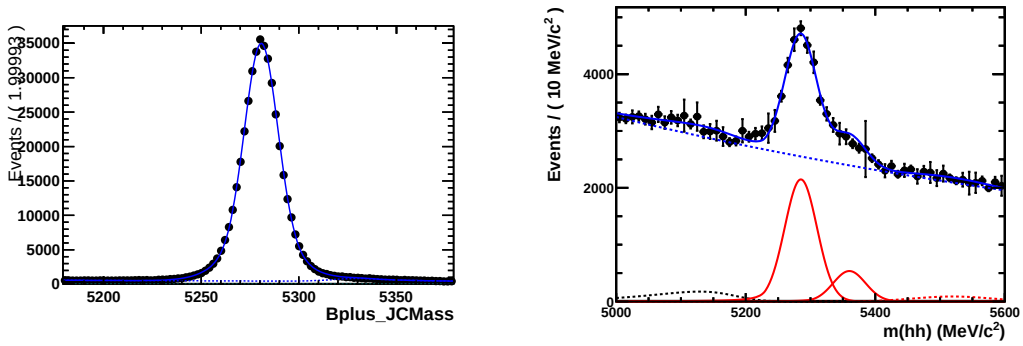
For the $B^0 \to K^+\pi^-$, the trigger efficiency is derived with events TIS only for L0 and HLT1, as requiring HLT2 TIS would reduce too much the sample. The HLT2 trigger efficiency is obtained with simulations and a systematic uncertainty is assigned. Plugging the trigger efficiency in Equation 3.3, the total number of $B^0 \to K^+\pi^-$ events simplifies, reducing further the uncertainties. Hence for this channel only the number of TIS $B^0 \to K^+\pi^-$ events are needed.

Control Channels' Event Yields

The last ingredient required to compute the normalisation factor defined in Equation 3.3 is the number of events observed in the control channels. These yields are extracted with a fit to the invariant mass spectrum of the $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$ samples. For $B^+ \to J/\psi K^+$, the model used to fit is composed of a double-Gaussian for the signal, an exponential for the combinatorial background and a Gaussian for the $B^+ \to J/\psi\pi^+$ component. $B^0 \to K^+\pi^-$ candidates are separated from $B_s^0 \to h^+h'^-$ with the method, based on the particle identification variables, used for the calibration (see Section 3.1.7). The invariant mass distributions of the two channels is shown in Figure 3.13. The obtained yields (with $1.0\,\text{fb}^{-1}$ of 2012 data) are 424200 ± 700 for $B^+ \to J/\psi K^+$ and the 14600 ± 1100 for the TIS $B^0 \to K^+\pi^-$ events at Ł0 and HLT1.

With these event yields the two normalisation factors can be computed for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$. They are then averaged accounting for their correlation. Table 3.4 shows that for each signal channel the two normalisation factors are compatible within their uncertainty and reports their averaged value.

With such normalisation factors, 13 $B_s^0 \to \mu^+\mu^-$ and 1.5 $B^0 \to \mu^+\mu^-$ events are expected for the SM for the $1.0\,\text{fb}^{-1}$of 2012 data in a 60 MeV mass window around the B_s^0 and the B^0 mass.



(a) (b)

Figure 3.13: Invariant mass distribution of (a) $B^+ \to J/\psi K^+$ and (b) $B^0_{(s)} \to h^+h'^-$ candidates.

Table 3.4: Normalisation factors for $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ for each normalisation channel and their resulting average for $1.0\,\text{fb}^{-1}$ of 2012 data.

Signal	$B^0 \to \mu^+\mu^-$	$B^0_s \to \mu^+\mu^-$
$\alpha_{B^+ \to J/\psi K^+}$	$(7.24 \pm 0.39) \times 10^{-11}$	$(2.83 \pm 0.27) \times 10^{-10}$
$\alpha_{B^0 \to K^+\pi^-}$	$(6.93 \pm 0.67) \times 10^{-11}$	$(2.71 \pm 0.34) \times 10^{-10}$
average	$(7.16 \pm 0.34) \times 10^{-11}$	$(2.80 \pm 0.25) \times 10^{-10}$

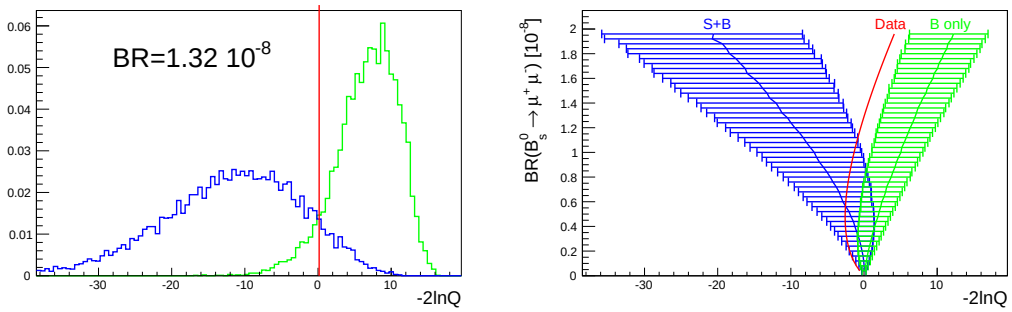
3.1.9 Limits and CL_s Method

Once the numbers of expected signal and background events have been derived, the observed data pattern can be compared to these expectations and if there is no clear evidence for signal, a range of branching fraction can be excluded. These upper limits are extracted with the CL_s method [24, 25]. The method relies on a scale on which observed data samples can be classified from the most background-like to the most signal-plus-background-like. This scale is usually referred to as the test statistic, noted Q:

$$Q = \prod \mathcal{P}(d_i, s_i + b_i)/\mathcal{P}(d_i, b_i) \ , \tag{3.6}$$

where $\mathcal{P}(d_i, s_i + b_i)$ is the probability that the expected number of events of signal and background $(s_i + b_i)$ has fluctuated according to a Poisson process to give the number of observed event (d_i) in the same bin.

Other definitions can be used for the test statistic but this one has the property that the joint test statistic for the outcome of two observed patterns



(a) (b)

Figure 3.14: (a) Distribution of the test statistic for the background only (B) hypothesis (green) and for the signal plus background (signal-plus-background) hypothesis (blue) for an assumed $\mathcal{B}(\mathrm{B_s^0} \to \mu^+\mu^-)$, the red line being an observed data pattern test statistic. (b) Medians and 1σ coverage of the background and signal-plus-background test statistic distribution as a function of $\mathcal{B}(\mathrm{B_s^0} \to \mu^+\mu^-)$. The red curve represents an observed data pattern test statistic as a function of $\mathcal{B}(\mathrm{B_s^0} \to \mu^+\mu^-)$.

is the product of the test statistic of the two observed patterns individually which is very convenient when combining experiments.

Once this scale has been defined, it has to be calibrated to know, for a given $\mathcal{B}(\mathrm{B_s^0} \to \mu^+\mu^-)$ hypothesis, which test statistic value is expected if the data being observed contain signal and background events and if they contain background only. This calibration is done by generating, for each $\mathcal{B}(\mathrm{B_s^0} \to \mu^+\mu^-)$ hypothesis, two sets of approximately 10k pseudo-experiments. The first one is generated by fluctuating the background only expectations according to statistical and systematic uncertainties while for the second one, signal events are also generated at a rate corresponding to the given $\mathcal{B}(\mathrm{B_s^0} \to \mu^+\mu^-)$ hypothesis. Each of the 10k pseudo-experiment has its test statistic computed. The test statistic distributions under these two hypotheses are shown in Figure 3.14(a) and their evolution as a function of the assumed $\mathcal{B}(\mathrm{B_s^0} \to \mu^+\mu^-)$ hypothesis is illustrated in Figure 3.14(b).

Once the scale has been calibrated, the distance of the observed data to the background only hypothesis and to the signal plus background hypothesis is quantified with the $\mathrm{CL_b}$ and $\mathrm{CL_{s+b}}$ as defined in Figure 3.15.

The $\mathrm{CL_b}$ allows to measure the compatibility of the data with the background hypothesis. Indeed, $1 - \mathrm{CL_b}$ gives the probability that only background produces the observed pattern. This quantity is usually referred as p-value. For an observed data sample, the lower this probability the more

$$CLb = \int_{Q<Q_{data}} Q_b \mathrm{d}Q_b$$

$$CLsb = \int_{Q<Q_{data}} Q_{sb} \mathrm{d}Q_{sb}$$

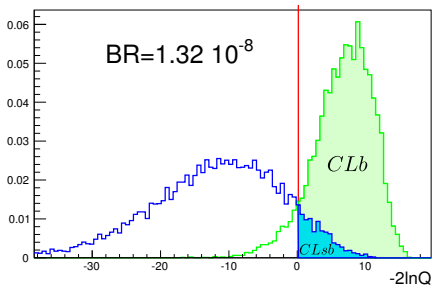
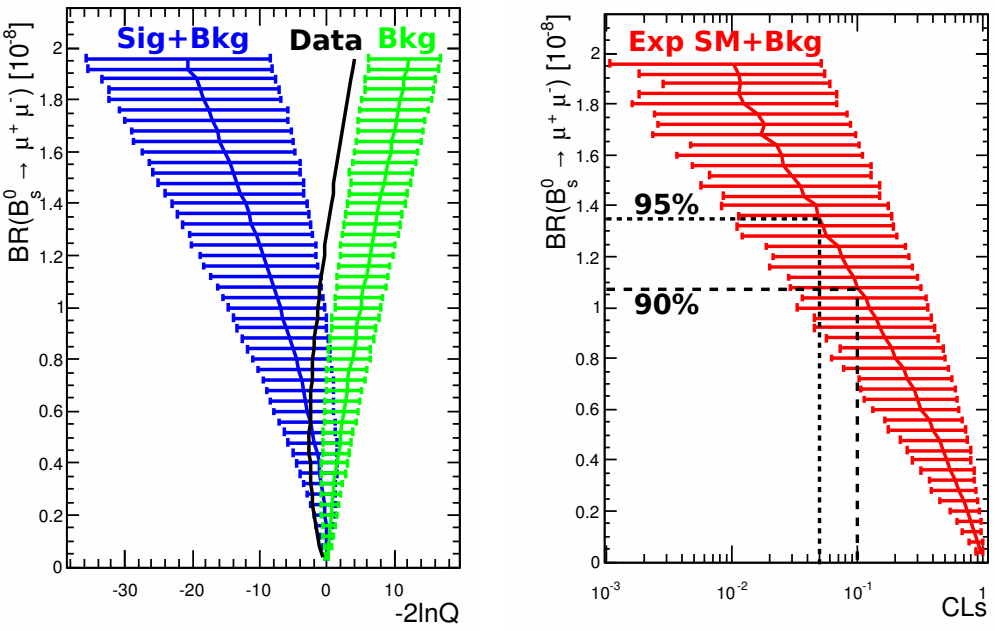


Figure 3.15: Definition of the $\mathrm{CL_{s+b}}$ and $\mathrm{CL_b}$. Q_{data} is the test statistic of the data and Q_b and Q_{sb} the test statistic under the background only hypothesis and the signal plus background hypothesis.

likely the dataset contains signal. A p-value of 2.7×10^{-3} indicates that to produce the observed pattern the background only would have fluctuated 3 standard deviations up from its central value, and indicates therefore the first evidence of signal. When this probability reaches 5.7×10^{-7}, the first observation of the signal at 5 standard deviations can be claimed. On the contrary, when the p-value is larger than these two thresholds, no signal evidence can be established but some branching fraction hypothesis can be excluded.

The compatibility of the data with the signal plus background hypothesis is evaluated by the $\mathrm{CL_s}$ defined as $\mathrm{CL_{s+b}}/\mathrm{CL_b}$. All the $\mathcal{B}(\mathrm{B_s^0} \to \mu^+\mu^-)$ hypothesis for which $\mathrm{CL_s}$ is larger than 0.05 (0.1) are excluded giving therefore an upper limit at 95% (90%) confidence level on the branching fraction. Figure 3.16(b) shows for each $\mathcal{B}(\mathrm{B_s^0} \to \mu^+\mu^-)$ the $\mathrm{CL_s}$ of the 2011 observed data [26] and the upper limits. Figure 3.16(a) depicts the evolution of observed data test statistic overlaid with the background only and the signal plus background hypotheses, and shows that all the $\mathcal{B}(\mathrm{B_s^0} \to \mu^+\mu^-)$ for which the data test statistic is far enough from the signal plus background test statistic distribution are excluded. Expected limits can be extracted by assuming that the test statistic of the observed data follows the central value of either the background only or the signal plus background test statistic distribution.

From Figure 3.16(a), the analysis sensitivity can be viewed as the ability to distinguish, from their test statistic values, the background only and the signal plus background hypotheses. If the signal plus background and background only test statistic distributions on Figure 3.14(a) were completely overlapping then the analysis would have no sensitivity to the branching fraction hypothesis.



(a) (b)

Figure 3.16: (a) Data test statistic compared to the expectations for the background-only and signal-plus-background hypotheses as a function of $\mathcal{B}(B^0_s \to \mu^+\mu^-)$. The error bars are centred at the median of the distribution and cover 34% on each side of the median. (b) CL_s as a functin of $\mathcal{B}(B^0_s \to \mu^+\mu^-)$. The 95% (90%) C.L. upper limit is obtained for $CL_s = 5\%$ (10%).

3.2 Binning Optimisation

The previous section explained that after the initial selection, the analysis strategy was to first classify events in bins of a multivariate discriminator and of the candidate invariant mass and then combine their results instead of applying an additional selection that would necessarily loose the sensitivity contained in the discarded events. This section describes how the bin boundaries of the two variables have been chosen in order to maximise the analysis sensitivity. First, Section 3.2.1 introduces a figure of merit to evaluate the analysis sensitivity with a given set of bins. Then, in order to compute this figure of merit for different sets of bins, a method is needed to evaluate the content of different bins without relying on statistical fluctuations. This method is described in Section 3.2.2. The results of the optimisation are shown in Section 3.2.3 and their performances are discussed in Section 3.2.4.

The binning optimisation is mainly constrained by the necessity to calibrate on data the BDT bins. This calibration method, described in Section 3.1.7, relies on fits of the $\mathsf{B}^0_{(\mathsf{s})} \to \mathsf{h}^+\mathsf{h}'^-$ sample and of the data invariant mass sidebands in each bin, therefore the bin boundaries must allow the two fits to be performed by assuring a minimum number of events. This binning optimisation process is documented in [27] and was originally developed for the analysis of the $1.0\,\text{fb}^{-1}$ of pp collisions recorded in 2011 [28]. The method was subsequently used by other LHCb analyses [29]. The results and the numbers presented in this section are derived based on the analysis of the $370\,\text{fb}^{-1}$ of pp collisions [26]. As far as the binning performances are concerned, the main difference between the latest analyses [28, 19] and the one from which the binning was optimised [26] is the tightening of the initial selection (see Section 3.1.5).

3.2.1 Quantifying the Sensitivity

The analysis sensitivity for a given branching fraction hypothesis can be seen as the separation of test statistic distributions generated under the signal plus background and background only hypotheses. An example of these distributions is shown in Figure 3.14(a). If the two distributions are well separated the observed data allow to exclude one of the two hypotheses on the basis of the distance of its test statistic to each distributions. Such a decision is harder or impossible when the two curves are overlapping. The analysis sensitivity can therefore be quantified, for a given branching fraction hypothesis with:

$$\Delta LQ \;=\; 2\ln Q_{SB}^{med} - 2\ln Q_{B}^{med}, \tag{3.7}$$

where Q_{SB}^{med} and Q_{B}^{med} are the median of the distribution of Q under the background only and the background plus signal hypotheses.

This quantity has also a concrete meaning when coarsely applying the Wilk's theorem [30]. Indeed, the ΔLQ can been seen as the χ^2 for the signal plus background distribution to be compatible with the background-only hypothesis, which translates directly as the squared signal significance. For example, a ΔLQ of 9.0 at the SM branching fraction of $\mathsf{B}_\mathsf{s}^\mathsf{0} \to \mu^+\mu^-$ would mean that a SM signal excess with a significance of 3 σ can be expected.

From the definition of the test statistic in Equation 3.6, Q_{SB}^{med} and Q_{B}^{med} can be evaluated without running any toys but using two proxies, $\widetilde{Q_{SB}^{med}}$ and $\widetilde{Q_{B}^{med}}$, defined by:

$$\widetilde{Q_{SB}^{med}} = \prod \mathcal{P}(s_i + b_i, s_i + b_i)/\mathcal{P}(s_i + b_i, b_i) \quad \text{i.e. data=sig+bkg} \tag{3.8}$$

$$\widetilde{Q_{B}^{med}} = \prod \mathcal{P}(b_i, s_i + b_i)/\mathcal{P}(b_i, b_i.) \quad \text{i.e. data=bkg} \tag{3.9}$$

Hence, the separation can be evaluated as:

$$\Delta LQ = 2ln\widetilde{Q_{SB}^{med}} - 2ln\widetilde{Q_{B}^{med}}. \tag{3.10}$$

The price for this huge CPU-time gain is that no-systematics errors can be accounted for in the optimisation. Nevertheless the effect of the systematics on the value of the distribution median are small. The accuracy of this evaluation was tested by comparing the test statistic median of the Q distribution under the background-only and signal-plus-background hypotheses obtained with 10k pseudo-experiments including systematics errors. Figure 3.17 shows good agreement between these medians and their evaluations from Equations 3.8 and 3.9.

These two evaluators provide an efficient tool to compare the separation of a great number (100k) of different binnings and therefore to selected the most optimal.

As explained in Section 3.1.7, each bin has to be calibrated. Hence the bin boundaries were taken from the nodes of a grid designed to ensure a minimum number of event in each sideband of the signal mass region. The order in which the two variables have their bin boundaries optimised is not important as these variables are not correlated. In practice the mass bin boundaries were optimised first.

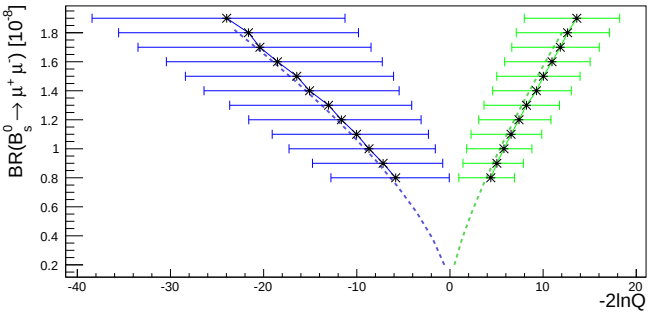


Figure 3.17: Central value and 1σ containment of the test statistic distributions generated for different branching fraction values for the signal plus background (blue) and background only (green) hypotheses. The dashed lines are the estimators of the test statistic central values of the two distributions.

3.2.2 Evaluation of the Expectation Signal and Background Bin Content

The signal and background bin contents used to evaluate the ΔLQ should be derived regardless of the statistical fluctuation of the data, otherwise a part of the sensitivity improvement would be artificial and the analysis results biased. The risk for such a bias is important as the data phase space driving the analysis sensitivity (i.e. the high BDT region) is by definition very poor in background events hence very sensitive to statistical fluctuations.

To avoid relying on these fluctuations, we chose to derive the background bin contents from a fit of the invariant mass and BDT distributions in the data mass sidebands. This technique, while keeping an accurate description of the data, smooths their statistical fluctuations and therefore avoids potential bias. The quality and stability of the fit was evaluated with Monte Carlo studies. The models used to fit was an exponential for the mass and the sum of three exponentials for the BDT:

$$\begin{aligned} f(m, bdt) &= \mathrm{N}_0 \times \mathrm{e}^{\mathrm{k}\cdot\mathrm{m}} \\ &\times \left(n_1 e^{sl_1 \cdot \mathrm{BDT}} + n_2 e^{sl_2 \cdot \mathrm{BDT}} + (1 - n_1 - n_2) e^{sl_3 \cdot \mathrm{BDT}}\right). \quad (3.11) \end{aligned}$$

Figure 3.18 shows the projections of the unbinned maximum likelihood fit performed using RooFit [31] and Figures 3.19, 3.20, and 3.22 the pulls of this fit after a toy MC study. The fit is found to be stable against the statistical fluctuations at high BDT.

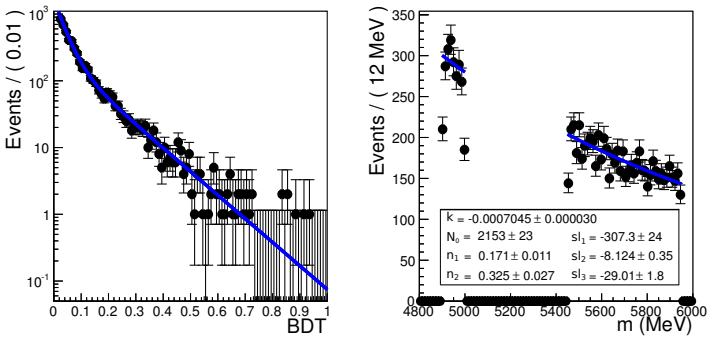


Figure 3.18: Two dimensional fit of the background mass sidebands $m\times$ BDT distribution. The three outlier points on the mass-fit plot are edge effects.

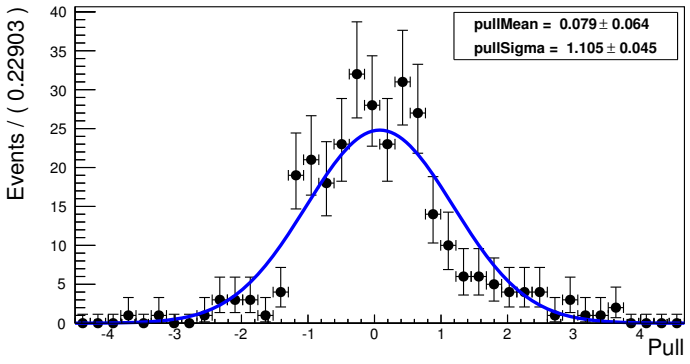


Figure 3.19: Pull of the slope $sl1$ of the first BDT exponential (see Equation 3.11).

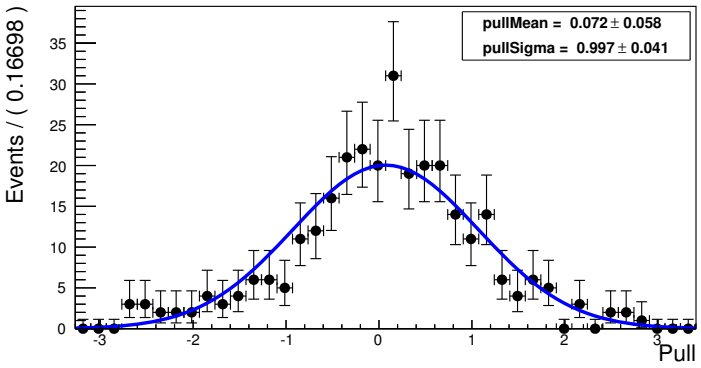


Figure 3.20: Pull of the slope $sl2$ of the second BDT exponential (see Equation 3.11).

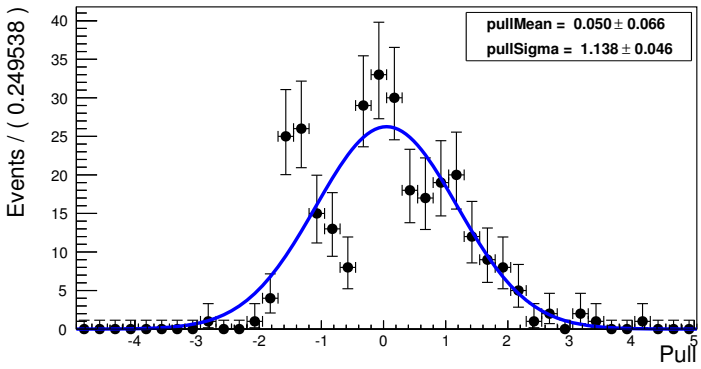


Figure 3.21: Pull of the slope $sl3$ of the third BDT exponential (see Equation 3.11).

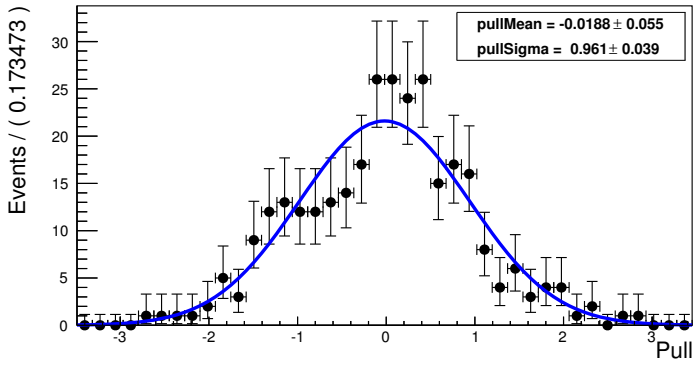


Figure 3.22: Pull of the slope of the mass exponential, k (see Equation 3.11).

For the signal expectation, no use of data was made. The signal PDF was assumed to be flat. The bin contents were derived by integrating this uniform PDF in the bin region. The fraction obtained was scaled by a coefficient depending on the normalisation factor and on the $\mathcal{B}(\mathrm{B}_s^0 \to \mu^+\mu^-)$ hypothesis, to get the final signal expectation.

3.2.3 Results

Mass Binning

The mass binning was optimised in the sensitive region BDT>0.8. Grid nodes were set every 6 MeV. The $\mathcal{B}(\mathrm{B}_s^0 \to \mu^+\mu^-)$ was set to the SM value $3.2 \cdot 10^{-9}$. All possible sets of 4 to 13 bins allowed by the grid were tested. For each number of bins, the bin set maximising the separation between the background-only and the signal plus background hypothesis was kept. Figure 3.24 shows how the separation evolves when the number of bins changes: the more bins, the more sensitive the analysis. However increasing the number of bins has a cost in term of complexity. Figure 3.24 shows the gain obtained by adding one more bin is significantly less important after nine bins. Hence we chose to use the 9-ranges binning shown on Figure 3.23. This binning is symmetric with respect to the mass central value and finer bins appear where the signal PDF has a larger derivative.

BDT Binning

Once the optimal mass binning was found, the BDT binning was optimised. Grid nodes were set every 0.05 but bins were imposed to have a minimum

Figure 3.23: Optimal mass binning.

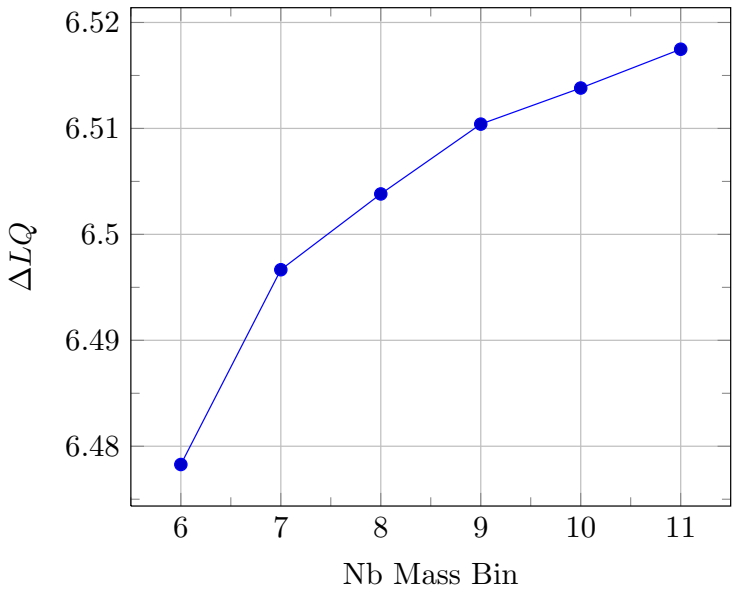


Figure 3.24: Separation as a function of the number of mass bins.

width of 0.1 in order to be calibrated. The $\mathcal{B}(\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-)$ was set to the SM value $3.2 \cdot 10^{-9}$. All possible sets of 5 to 9 bins allowed by this grid and matching the minimal width constraints were tested. For each number of bins the set maximising the separation between the background only and signal plus background hypotheses was kept. Figure 3.25 shows that the separation increases with the number of bins and tends to reach a plateau. The saturating separation value, S_{max}, was extracted by fitting the points with $S_{max}(1 - e^{-n_{bin}/n_0})$. The fit result is shown in Figure 3.26. We decided to keep the set of 8 bins, shown in Figure 3.27, as it reaches a separation larger than 99% of the plateau value. The contents of the bins derived from the two-dimensional fit (integrating over the mass window) are reported in Table 3.5.

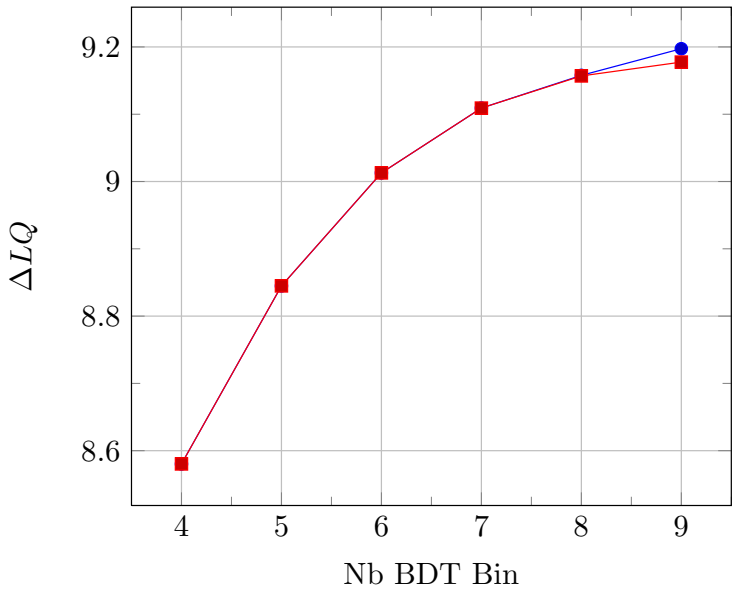


Figure 3.25: Separation as a function of the number of BDT bins. For the red (squared markers) curve, the binnings have a minimum bin width of 0.1, while for the blue (circle markers) no constrain on the bin width is imposed.

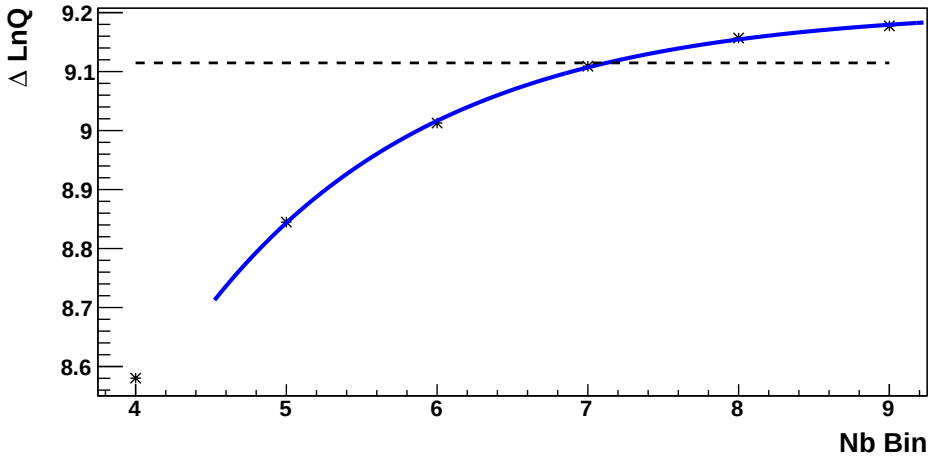


Figure 3.26: Exponential fit of the ΔLQ vs number of BDT bins points (the minimum bin width is constrained to 0.1). The horizontal dashed line is set at 99% of the maximum ΔLQ according to the fit.

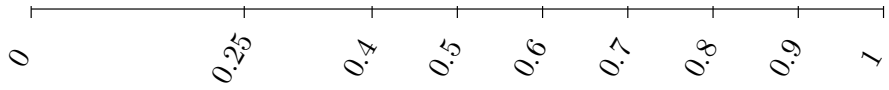


Figure 3.27: Optimal BDT binning with a minimal bin width of 0.1.

	0-0.25	0.25-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1
sig	2.581	1.548	1.032	1.032	1.032	1.032	1.032	1.032
bkg	2061	65.45	15.11	6.704	2.975	1.320	0.5859	0.2600

Table 3.5: BDT bin content obtained from the two-dimensional fit to the $370\,\mathrm{pb}^{-1}$ data mass sideband.

3.2.4 Discussion of the Performances

In order to quantify the sensitivity gain brought by the new binning the ΔLQ was computed with the old sets (4 regular bins in BDT and 6 regular one in mass). The results of these computations are reported in Table 3.6. The new binning brings a 14% improvement, raising the ΔLQ from 8.04 to 9.16. The improvement mainly comes from the BDT binning. If the Wilk's theorem can be applied, this new binning allows to pass the critical value of $\Delta LQ = 9$, which corresponds to an evidence with a significance of 3 σ.

Mass Binning	BDT Binning	ΔLQ
New 9 bins	New 8 bins	9.16
Old 6 bins	New 8 bins	9.07
New 9 bins	Old 4 bins	8.12
Old 6 bins	Old 4 bins	8.04

Table 3.6: ΔLQ of the new binning (first row) compared (3 last rows) to the ΔLQ obtained when an old binning (BDT or mass) is used.

The contribution of each bin to the analysis sensitivity can be evaluated using the additivity of the ΔLQ:

$$\Delta LQ = \sum_{bin} \Delta LQ_i,$$

where ΔLQ_i is the ΔLQ of bin i. Figure 3.28 shows the contributions that each bin brings to the total ΔLQ and their ranking with respect to this

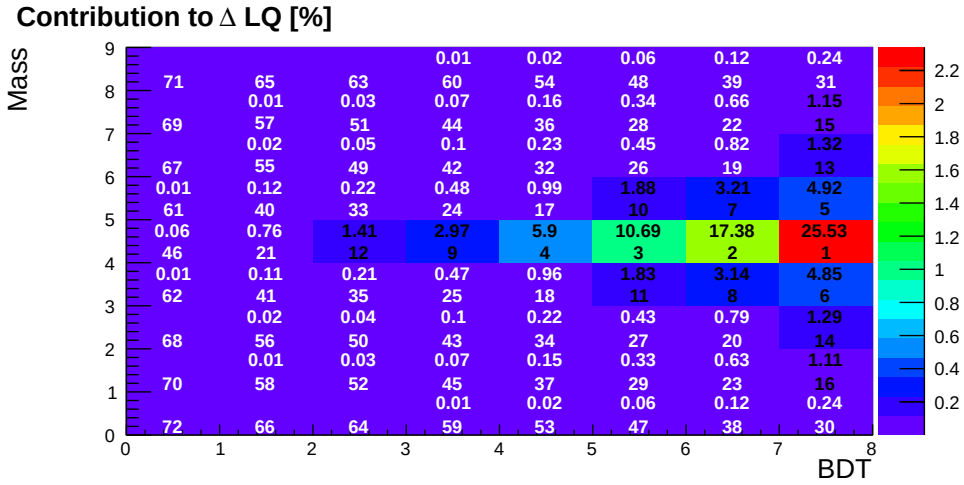


Figure 3.28: The 9×8 bins were ranked by decreasing order of ΔLQ_i. The top label in each bin is the fraction that this ΔLQ_i represents with respect to the total ΔLQ (when this label is missed, the ΔLQ_i is smaller than 0.01%). The bottom label is the rank of the bin.

criterium. Table 3.7 presents the contributions when bins are grouped by BDT.

The first BDT bin appears to contribute to less than 0.1% of the analysis sensitivity. This conclusion, in light of the difficulty raised by the calibration of the first bin due to the poor $B^0_{(s)} \to h^+h'^-$ signal significance, contributed to review the analysis selection for the following analysis [28] and introduce a tighter multivariate selection removing a large fraction of this first BDT bin. Figure 3.29 shows how the analysis ΔLQ evolves when bins are added from the most to the least sensitive one.

3.3 Upper Limits and p-value Extraction

This section describes the extraction of the limits and p-values with the CL_s method for the 2012 analysis [19]. The general structure of the implementation of this method is described in Section 3.3.1. The treatment of the statistical and systematic uncertainties is explained in Section 3.3.2. The computation of the test statistic is presented in Section 3.3.3. Finally sensitivity projections based on the previous analysis [28] are described in Section 3.3.4. Note that the observed results are presented only in the next section.

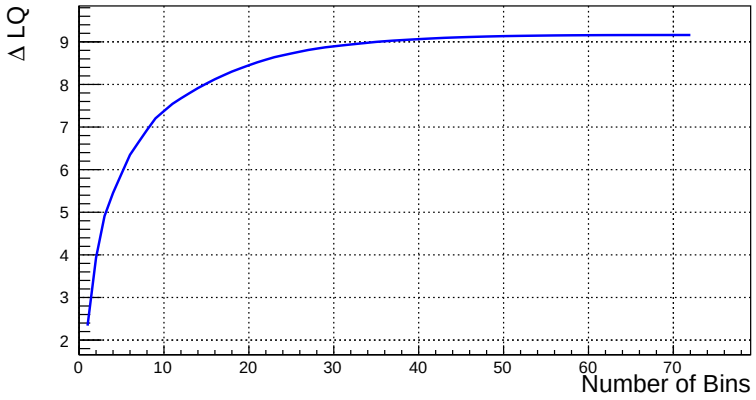


Figure 3.29: Evolution of the ΔLQ when bins are added by decreasing order of contribution to the analysis sensitivity.

	BDT bin i	ΔLQ_i	$\Delta LQ_i/\Delta LQ$ (%)
1:	0 – 0.25	0.00899	0.0981
2:	0.25 – 0.4	0.100	1.09
3:	0.4 – 0.5	0.186	2.03
4:	0.5 – 0.6	0.396	4.33
5:	0.6 – 0.7	0.797	8.70
6:	0.7 – 0.8	1.48	16.1
7:	0.8 – 0.9	2.47	26.9
8:	0.9 – 1	3.73	40.7

Table 3.7: The ΔLQ_i of each BDT bin i are presented in the second column. The fraction each ΔLQ_i represents with respect to the total ΔLQ is shown in the third column.

3.3.1 Implementation of the CL_s Method

The main ideas of the CL_s method have been described in Section 3.1.9. Among the steps of the method, the calibration of the test statistic with pseudo experiments for the different branching fraction hypotheses is the most challenging one to implement technically. Indeed, for the method to provide accurate results, the pseudo experiments must be generated and treated, as much as possible, as the observed data. The level at which the analysis has to be reproduced with the pseudo-experiments is roughly the one at which the analysis has been described in Section 3.1.

For each pseudo-experiment, the data pattern that would be observed is generated with a list of parameters such as, for example, the central value of the signal invariant mass distribution. These parameters must be decorrelated one from another which requires, for the 2012 analysis, more than 150 parameters to accurately describe a pseudo-experiment. Each of these parameters is associated to a prior PDF describing its uncertainties and implemented with the RooFit package [31]. This technical solution led to a major progress with respect to the standard code usually referred as MCLimits [32] which was used before. Indeed the new method allows to describe analytically any PDF while this was impossible with the MCLimits.

The pseudo-experiment generation process is implemented in the RootStat [33] package with a C++ class based on the RooHybridCalculator. The first step of the process is to generate all parameters according to their PDF. This set of values defines the signal and background PDF as well as the expected SM signal and background yields. These expected yields define Poisson distributions which are used to generate the actual number of events for the pseudo experiment. The technical details of this data pattern production are described in details in Section 3.3.2. After these three data patterns (signal, SM signal, and background) are produced, the signal and background PDFs and the events yields are restored to their initial values and used as reference to compute the test statistic of the generated data patterns. The derivation of the input parameters' means is described more precisely in Section 3.3.3. All parameters generated during this process as well as the three test statistic values are stored in a n-tuple. From these records the distributions of the number of SM and background events can be extracted, as well as the test statistic distributions. The former are used to derive the number of expected events of each species in each BDT bin while the three latter are used to compute CL_{s+b} and CL_b as defined in Section 3.1.9.

The CL_{s+b} (CL_b) of the observed data pattern is obtained for each branching fraction hypothesis by deriving the proportion of pseudo-experiments generated under the signal plus background (background only) hypothesis

which have a test statistic larger than that of the observed data pattern. The CL_s versus branching fraction curve for this observed data pattern is obtained by taking the ratio of the CL_{s+b} to the CL_b for each branching fraction hypothesis.

By performing the same procedure with every SM-signal plus background and background only pseudo-experiment used as observed data pattern, distributions of CL_{s+b} and CL_b and hence CL_s are built for each branching fraction hypothesis. The expected CL_s versus branching fraction curves under the signal plus background and background only hypotheses are obtained by extracting for each branching fraction the distribution median. The spread of the distribution is reported with error bars covering 68% of the distribution and defined by the two points for which 16% of the distribution lies above or below.

3.3.2 Pseudo-experiment Generation

The data patterns generated for each pseudo-experiment are made of four components: the signal, the combinatorial background, the misidentified $\mathsf{B}^0_{(s)} \to \mathsf{h}^+\mathsf{h}'^-$ background, and the cross-feed of the signals. As quickly mentioned in the previous section, these four components are generated from a list of basic and uncorrelated input parameters via a process accounting for the systematic uncertainties by fluctuating the input parameters and, for the statistical ones, by generating the total number of event of each species according to a Poissonian of mean set to the expected numbers of events derived from the fluctuated input parameters. We describe here how for each species the number of expected events in the BDT and invariant mass are derived.

Signals and Cross-feeds

The $\mathsf{B}^0_s \to \mu^+\mu^-$ and $\mathsf{B}^0 \to \mu^+\mu^-$ invariant mass PDFs are Crystal Ball functions parametrised by their means, resolutions, and transition points. The first two are fluctuated according to a Gaussian PDF, while the latter is fixed. From these two PDFs, each mass bin content is obtained by integrating over its bin range and scaling to the PDF integral in the signal mass window.

The yield is obtained by dividing the branching fraction hypothesis (and the SM one for the cross-feed) by the average α_{norm}, defined in Equation 3.3. This average normalisation factor is in fact derived from more elementary uncorrelated parameters since $f_{\text{norm}}/f_{\mathsf{B}^0_{(s)}}$ enter both $\alpha_{\mathsf{B}^0\to\mathsf{K}^+\pi^-}$ and $\alpha_{\mathsf{B}^+\to\mathsf{J}/\psi\mathsf{K}^+}$, and the total number of TIS $\mathsf{B}^0 \to \mathsf{K}^+\pi^-$ is also used to derived total number

of misidentified $B^0_{(s)} \to h^+h'^-$. All these elementary parameters are fluctuated according to Gaussian PDFs.

The BDT PDF is obtained from the proportions of $\mathrm{TIS}_{\mathrm{L0,HLT1}}$ $B^0_{(s)} \to h^+h'^-$ in each bin corrected by factors accounting for the HLT2 trigger efficiency and for the signal to control channels trigger efficiency ratio:

$$\frac{\epsilon_{\mathrm{L0,HLT1,HLT2}}(B^0_s \to \mu^+\mu^-)}{\epsilon^{\mathrm{TIS}}_{\mathrm{L0,HLT1}}(B^0_{(s)} \to h^+h'^-) \times \epsilon_{\mathrm{HLT2}}}.$$

$B^0_{(s)} \to h^+h'^-$ Misidentified Background

The $B^0_{(s)} \to h^+h'^-$ misidentified background originates mainly from $B^0_{(s)} \to h^+h'^-$ where the two hadrons decay in flight to muons. The mass PDF of this component is obtained by smearing the hadron tracks momenta in data in order to replicate the momentum shift observed in simulation. The smearing magnitude is treated as a systematic uncertainty. A larger smearing spreads more the initial distribution so bin contents in the sides of the distribution increase while those in the central part decrease and vice versa for a smaller smearing as shown in Figure 3.30. This systematic uncertainty is implemented by recomputing the bin content after the smearing magnitude is fluctuated and applied.

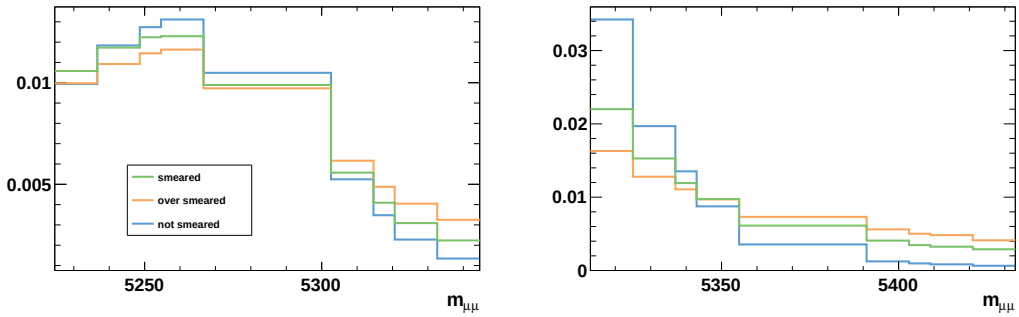
The BDT PDF is obtained from the number of TIS $B^0_{(s)} \to h^+h'^-$ in each bin[7]. This PDF is then corrected by the double misidentification probability evaluated in bins of momentum and transverse momentum with a control sample.

Finally, the event yield is obtained from:

- the number of TIS $B^0 \to K^+\pi^-$ in the two signal mass windows,
- the product of the misidentification probabilities and the $B^0_{(s)} \to h^+h'^-$ to TIS $B^0 \to K^+\pi^-$ efficiency ratio,
- and the fraction of events falling in each signal mass window.

Here again great care was taken to account for the correlations with the $B^0_{(s)} \to \mu^+\mu^-$ BDT and with all quantities using the number of TIS $B^0 \to K^+\pi^-$.

[7]The correction due to the time acceptance, described in Section 4, should also not be applied to the $B^0_{(s)} \to h^+h'^-$ BDT PDF.



(a) (b)

Figure 3.30: Invariant mass shape in the $B^0 \to \mu^+\mu^-$ (a) and $B_s^0 \to \mu^+\mu^-$ (b) mass windows obtained from the simulated data by smearing both hadron tracks. The green line is the shape obtained with the normal smearing, the orange one with a smearing as large as for K applied to π and the blue one with no smearing and just the wrong mass hypothesis.

Combinatorial Background

The mass PDF of the combinatorial background is modelled with a single exponential function in each BDT bin and a systematic uncertainty accounting for the arbitrariness of the model choice is set by comparing with the results obtained using a double exponential function.

The BDT bin contents are obtained from the slope of the exponential function, the numbers of events in the data invariant mass sidebands and the fraction of these events that are combinatorial backgrounds, as the sidebands also contain contributions from $B^{0,+} \to \pi^{0,+}\mu^+\mu^-$ and $B^0 \to \pi^-\mu^+\bar{\nu}_\mu$. These fractions are obtained with their uncertainties by applying the misidentification probability to simulated data and normalising to $B^+ \to J/\psi K^+$ to get the yields.

The number of events in the mass sidebands are very small for the high BDT bins and their fluctuations are therefore generated according to a Poissonian distribution. This type of distribution is defined by its mean but the quantities observed in the data mass sidebands are the most probable values of the distribution and at low statistics the two quantities are different. Following [34], the mean of the Poisson distribution, $\langle N \rangle$ was evaluated from its

most probable value, $\hat{N}$, by:

$$\begin{aligned} \langle N \rangle &= \hat{N} \exp\left(\frac{1}{2\hat{N}} - \frac{1}{12\hat{N}^2}\right) \\ &\simeq \hat{N} + 0.5. \end{aligned} \tag{3.12}$$

3.3.3 Test Statistic Computation

Once a data pattern, (i.e. a set of numbers of events for each bin, d_i) is generated, its test statistic is computed using Equation 3.6 where the probabilities $\mathcal{P}(d_i, s_i+b_i)$ and $\mathcal{P}(d_i, b_i)$ assume Poissonian processes. For this computation to be correct, the expected numbers of signal and background events, s_i and b_i, should correspond to the means of the Poissonian distributions. In the case of the background, b_i are obtained by interpolating the number of events observed in the sidebands with an exponential function. These numbers of events are obviously the most probable values of Poissonian processes hence we should be careful and check that the function used to interpolate returns the mean of background distributions in the mass windows and not their most probable value. The three simplest functions that can be used for this interpolation are:

$$\begin{aligned} b_i^A &= \hat{N}_i \frac{\int_{\mathrm{bin_i}} e^{-k_i m} \mathrm{d}m}{\int_{\mathrm{SB_i}} e^{-k_i m} \mathrm{d}m}, \\ b_i^B &= \langle N_i \rangle \frac{\int_{\mathrm{bin_i}} e^{-k_i m} \mathrm{d}m}{\int_{\mathrm{SB_i}} e^{-k_i m} \mathrm{d}m}, \\ b_i^C &= \langle \hat{N}_i \frac{\int_{\mathrm{bin_i}} e^{-k_i m} \mathrm{d}m}{\int_{\mathrm{SB_i}} e^{-k_i m} \mathrm{d}m} \rangle, \end{aligned}$$

where $\hat{N}_i$ is the number of events observed in the data sideband i, the chevron signs $\langle\rangle$ indicates that the value inside the signs is assumed to be the most probable value of a Poissonian distribution and is converted into its mean according to Equation 3.13. k_i is the slope of the exponential function modelling the invariant mass distribution in the BDT bin i. These three estimators were tested by generating for each BDT bin the numbers of observed events in the sidebands, N_i^{gen}, according to Poisson distributions of means $\langle N_i \rangle$ and deriving the bin content by:

$$b_i^A = N_i^{gen} \frac{\int_{\mathrm{bin_i}} e^{-k_i m} \mathrm{d}m}{\int_{\mathrm{SB_i}} e^{-k_i m} \mathrm{d}m}.$$

Then the mean of the distribution built with this procedure was compared to the value returned by the three estimators. The first estimators, b_i^A, was found to return a value at least two times closer to the mean than the others and was therefore chosen to derive the combinatorial background expectations.

3.3.4 Sensitivity Projection

Expected upper limits can be extrapolated to larger integrated luminosity values by scaling the signal and background expectations. This was done after each analysis, but we only present here the projections based on the $1\,\text{fb}^{-1}$ analysis [28]. For this projection the scaling must also account for the increase of the pp centre of mass energy as the $\mathsf{b\bar{b}}$ production cross section raises approximately linearly with it as shown in Figure 2.4. The scaling process should also include systematic uncertainties, nevertheless achieving a realistic projection of these uncertainties is more difficult as some parameters like f_s/f_d are inputs to the analysis. On the other hand, the analysis is still largely dominated by the statistical uncertainties hence, the systematic uncertainties were not accounted for in the projection.

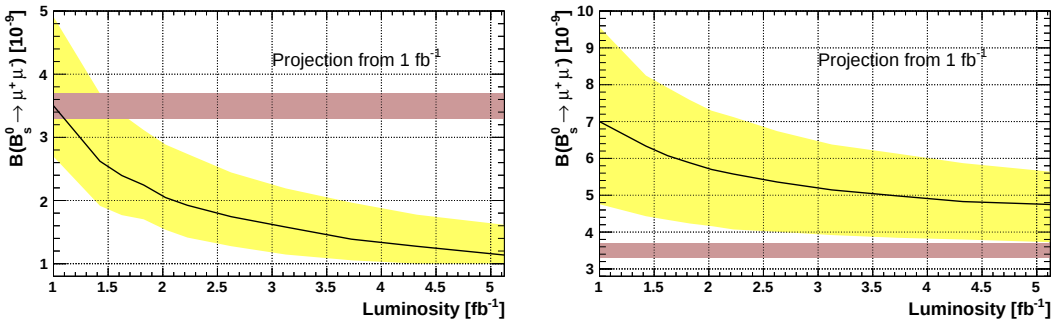
This projection is obtained by building, for several integrated luminosity hypotheses, the CL_s curves under the background only and signal plus background hypothesis as a function of the branching fraction. Then, for each of these curves, the median 95% C.L. upper limits and their 68% C.L. interval are extracted. Figure 3.31 reports the expected upper limits as a function of the integrated luminosity.

The discovery potential of the analysis can also be extrapolated by building for each integrated luminosity the CL_b curve as a function of the $\mathsf{B^0_s} \to \mu^+\mu^-$ branching fraction and extracting the branching fractions for which the median CL_b value corresponds to an evidence with a significance of 3σ ($\text{CL}_\text{b} = 1 - 2.7 \times 10^{-3}$). Figure 3.32 gives an example of such a CL_b curve and Figure 3.33 shows the resulting projection curve.

A more realistic discovery potential projection can be obtained by including the knowledge brought by the already un-blinded data set. A simultaneous fit to the invariant mass and BDT distributions of these data indicates that the $\mathsf{B^0_s} \to \mu^+\mu^-$ signal was most likely produced at a rate of $0.8^{+1.8}_{-1.3} \times 10^{-9}$. Therefore to reach the number of signal events required to have an evidence with a significance of 3σ, which was given by:

$$N_{sig}^{3\sigma} \;\propto\; \mathcal{B}(\mathsf{B^0_s} \to \mu^+\mu^-)^{3\sigma} + \frac{8}{7}(1-\mathcal{L}) \times \mathcal{B}(\mathsf{B^0_s} \to \mu^+\mu^-)^{3\sigma}, \quad (3.13)$$

the signal events must be produced in the new data at a higher rate, referred



(a) (b)

Figure 3.31: 95% C.L. expected upper limit on the $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ branching fraction as a function of the integrated luminosity if background only was observed (a) and if the signal SM and background were observed (b). The black line is the $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ branching fraction for which the $\mathrm{CL_s}$ distribution median is 5%. The yellow area covers all branching fraction values which could be excluded if the $\mathrm{CL_s}$ fluctuated within one standard deviation from its median value. The brown horizontal thick line is the time integrated branching fraction prediction for the SM.

to as $\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-)^{3\sigma}_{new}$ such that:

$$N^{3\sigma}_{sig} \quad \propto \quad 0.8 + \frac{8}{7}(1-\mathcal{L}) \times \mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-)^{3\sigma}_{new}. \tag{3.14}$$

In the previous expressions the $\frac{8}{7}(1-\mathcal{L})$ factors arises to account for the increase, after $\mathcal{L} = 1.0\,\mathrm{fb}^{-1}$, of the energy centre of mass from 7 TeV up to 8 TeV.

Figure 3.34 shows $\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-)^{3\sigma}_{new}$ as a function of the integrated luminosity. At 68% C.L. under the assumption of the SM, the first evidence of $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ is observed with an integrated luminosity between 1.7 and 3.3 fb^{-1}.

Finally the probability of having an evidence with a significance larger than 3σ for SM signal and at a given integrated luminosity, can be estimated by generating pseudo-experiments, with the SM rate, and deriving the fraction of pseudo-experiments reaching a $\mathrm{CL_b}$ larger than the value corresponding to an evidence with a significance of 3σ $(1-2.3\times10^{-3})$. The $\mathrm{CL_b}$ distribution for an integrated luminosity of around $2\,\mathrm{fb}^{-1}$ is shown on Figure 3.35(a) and the probability of having an evidence for the SM as a function of the luminosity on Figure 3.35(b). The probability of having an evidence with a significance larger than 3σ for SM signal with $2.1\,\mathrm{fb}^{-1}$ of integrated luminosity is 68% and the projected p-value is 4×10^{-4}. However this figure

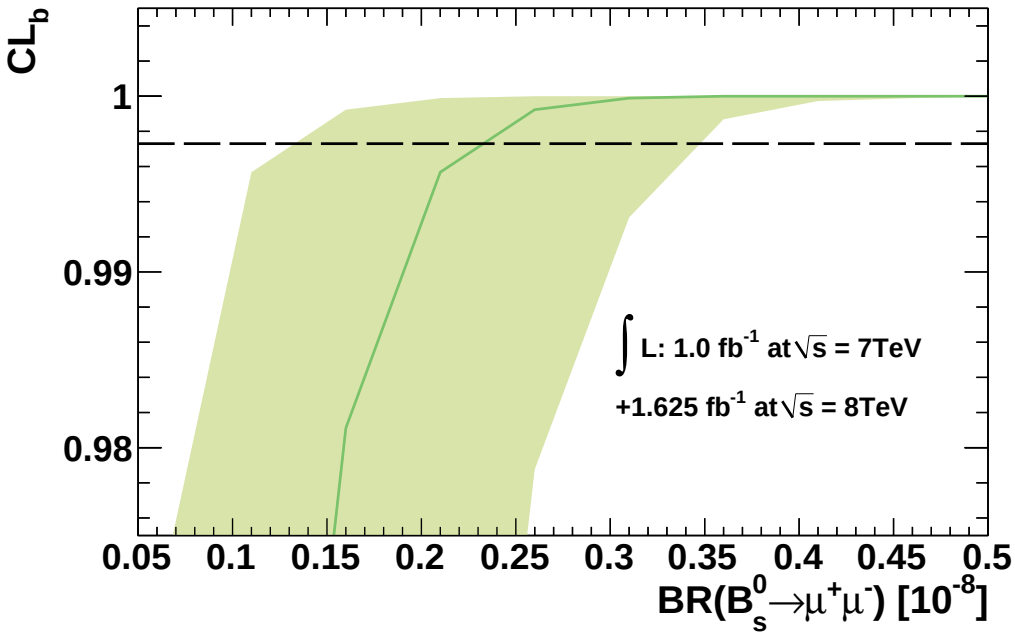


Figure 3.32: CL_b as a function of the $B_s^0 \to \mu^+\mu^-$ branching fraction hypothesis. The green line is the median of the CL_b distributions obtained for each branching fraction hypothesis and the light green area covers 34% on these distributions on each side of their median. The horizontal dashed line is the CL_b value corresponding to an evidence with a significance of 3σ.

does not account for the data pattern observed in 2011 in which the signal yield was less than what expected in the SM and is therefore overestimating this probability.

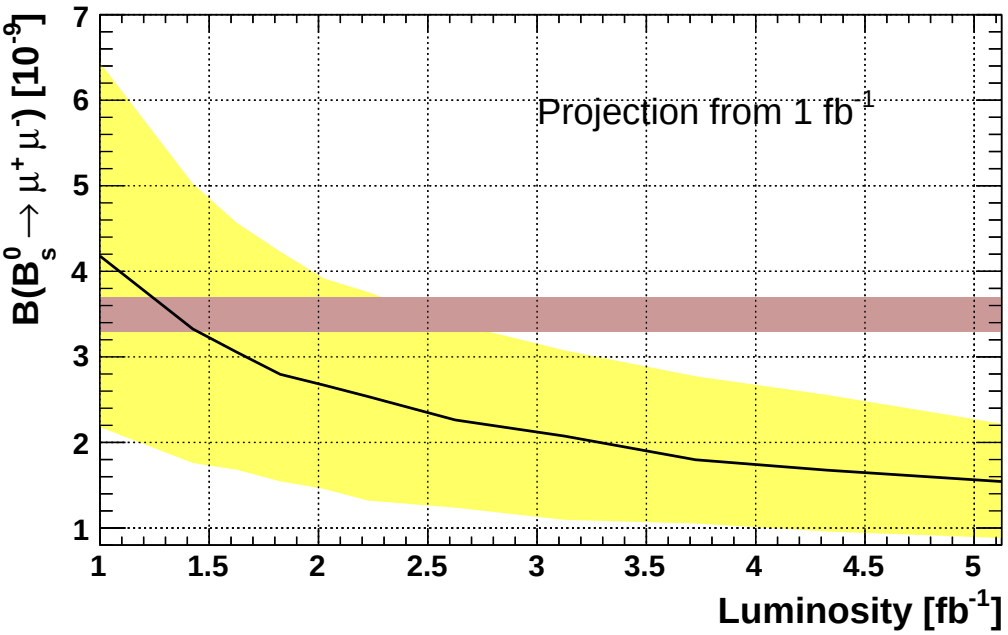


Figure 3.33: $B^0_s \to \mu^+\mu^-$ branching fraction at which an evidence with a significance of 3σ could be reached as a function of the integrated luminosity. Observed data pattern in the first $1.0\,\text{fb}^{-1}$ are not accounted for. The black line is the $B^0_s \to \mu^+\mu^-$ branching fraction for which the CL_b distribution median corresponds to a p-value of 2.7×10^{-9}. The yellow area covers all branching fraction values for which an evidence with a significance of 3σ could be observed if the CL_b fluctuated within one standard deviation from its median value. The red horizontal area is the time integrated branching fraction prediction for the SM.

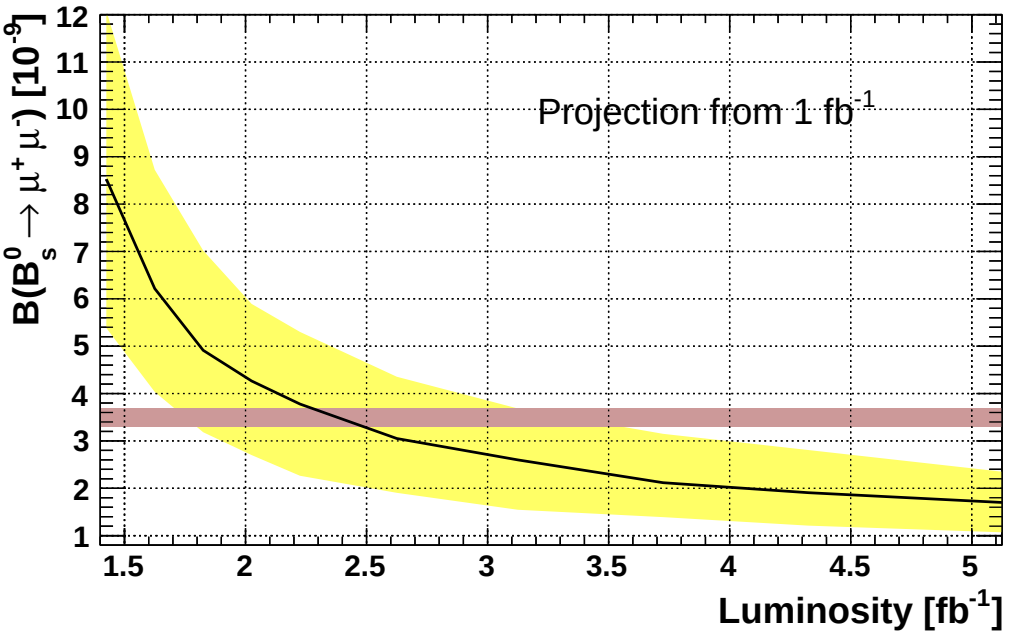
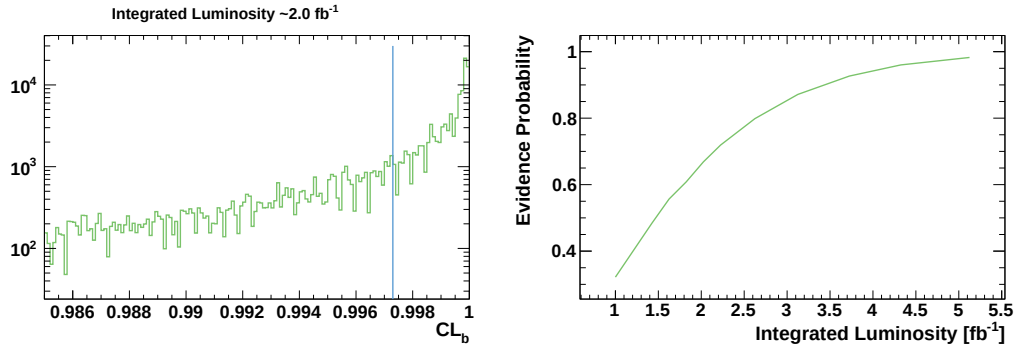


Figure 3.34: $B_s^0 \to \mu^+\mu^-$ branching fraction at which an evidence with a significance of 3σ could be reached as a function of the integrated luminosity. Observed data pattern in the first $1.0\,\mathrm{fb}^{-1}$ are accounted for. The black line is the $B_s^0 \to \mu^+\mu^-$ branching fraction for which the CL_b distribution median corresponds to a p-value of 2.7×10^{-3}. The yellow area covers all branching fraction values for which an evidence with a significance of 3σ could be observed if the CL_b fluctuated within one standard deviation from its median value. The brown horizontal area is the time integrated branching fraction prediction for the SM.



(a) (b)

Figure 3.35: (a) CL_b distribution for pseudo-experiments generated under the assumption of a SM signal and an integrated luminosity of $2\,fb^{-1}$. The vertical blue line is the CL_b value corresponding to an evidence with a significance of 3σ. (b) Probability of having an evidence with a significance larger than 3σ as a function of the integrated luminosity.

Table 3.8: Expected combinatorial background, $B^0_{(s)} \to h^+h'^-$ background, cross-feed, and signal events assuming SM predictions, together with the number of observed events in the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ mass signal regions, in bins of BDT for the 2012 data sample.

Mode	BDT bin	0.0 – 0.25	0.25 – 0.4	0.4 – 0.5	0.5 – 0.6	0.6 – 0.7	0.7 – 0.8	0.8 – 1.0
$B^0_s \to \mu^+\mu^-$	Exp. comb. bkg	2345^{+40}_{-40}	$56.7^{+3.0}_{-2.9}$	$13.1^{+1.5}_{-1.4}$	$4.42^{+0.91}_{-0.81}$	$2.10^{+0.67}_{-0.56}$	$0.35^{+0.42}_{-0.22}$	$0.39^{+0.33}_{-0.21}$
	Exp. peak. bkg	$0.250^{+0.083}_{-0.068}$	$0.145^{+0.049}_{-0.040}$	$0.081^{+0.027}_{-0.023}$	$0.075^{+0.024}_{-0.020}$	$0.071^{+0.023}_{-0.019}$	$0.062^{+0.021}_{-0.017}$	$0.104^{+0.034}_{-0.028}$
	Exp. signal	$3.69^{+0.59}_{-0.52}$	$2.14^{+0.37}_{-0.33}$	$1.20^{+0.21}_{-0.18}$	$1.16^{+0.18}_{-0.16}$	$1.17^{+0.18}_{-0.16}$	$1.15^{+0.19}_{-0.17}$	$2.13^{+0.33}_{-0.29}$
	Observation	2274	65	19	5	3	1	3
$B^0 \to \mu^+\mu^-$	Exp. comb. bkg	2491^{+42}_{-42}	$59.5^{+3.3}_{-3.2}$	$13.9^{+1.6}_{-1.5}$	$4.74^{+1.00}_{-0.89}$	$2.10^{+0.74}_{-0.61}$	$0.55^{+0.50}_{-0.31}$	$0.29^{+0.34}_{-0.19}$
	Exp. peak. bkg	$1.49^{+0.50}_{-0.36}$	$0.86^{+0.29}_{-0.22}$	$0.48^{+0.16}_{-0.12}$	$0.44^{+0.15}_{-0.11}$	$0.42^{+0.14}_{-0.10}$	$0.369^{+0.126}_{-0.093}$	$0.62^{+0.21}_{-0.15}$
	Exp. cross-feed	$0.627^{+0.104}_{-0.091}$	$0.363^{+0.066}_{-0.057}$	$0.204^{+0.036}_{-0.032}$	$0.197^{+0.032}_{-0.027}$	$0.199^{+0.032}_{-0.028}$	$0.196^{+0.034}_{-0.030}$	$0.362^{+0.058}_{-0.051}$
	Exp. signal	$0.442^{+0.062}_{-0.057}$	$0.256^{+0.040}_{-0.036}$	$0.144^{+0.022}_{-0.020}$	$0.139^{+0.019}_{-0.017}$	$0.140^{+0.019}_{-0.018}$	$0.138^{+0.021}_{-0.019}$	$0.255^{+0.035}_{-0.031}$
	Observation	2433	59	19	3	2	2	2

3.4 First Evidence of $B^0_s \to \mu^+\mu^-$

This section describes the results of the 2012 analysis [19]. This analysis combines the results obtained with the $1.1\,\text{fb}^{-1}$ of integrated luminosity collected in 2012 together with the 2011 sample ($1.0\,\text{fb}^{-1}$) which have been re-analysed with a more refined background parametrisation with respect to [28]. With the background parametrisation update, the two analyses are essentially identical apart for the signal BDT calibration which uses in 2012 particle identification to distinguish the different $B^0_{(s)} \to h^+h'^-$ modes. The 2011 analysis statistics in the sensitive region allows to calibrate 8 BDT bins while the two most sensitive bins have to be merged in the 2012 analysis to achieve their calibration. The expected results of the individual and combined analyses are described in Section 3.4.1 and are then compared in Section 3.4.2 with the observed ones. A $B^0_s \to \mu^+\mu^-$ signal evidence being observed, the decay branching fraction can be measured with a fit whose results are given in Section 3.4.3.

3.4.1 Expectations

If the signal rate is the one predicted by the SM, the projections described in Section 3.3.4 foresee a p-value of 4×10^{-4} and indicate that an evidence with a significance larger than 3 σ has a 68% probability to be observed. Lacking an observation, an expected upper limits at around 6.6×10^{-9} could be set for $\mathcal{B}(B^0_s \to \mu^+\mu^-)$.

Using 2012 data sidebands these projections can be refined. This additional knowledge is summarised by the first three rows of Table 3.8 and Table 3.9 which report the expected signal and backgrounds events obtained with the CL_s code when SM pseudo-experiments are generated. These pseudo-experiments allow to compute the compatibility of the SM data patterns with the background only hypothesis.

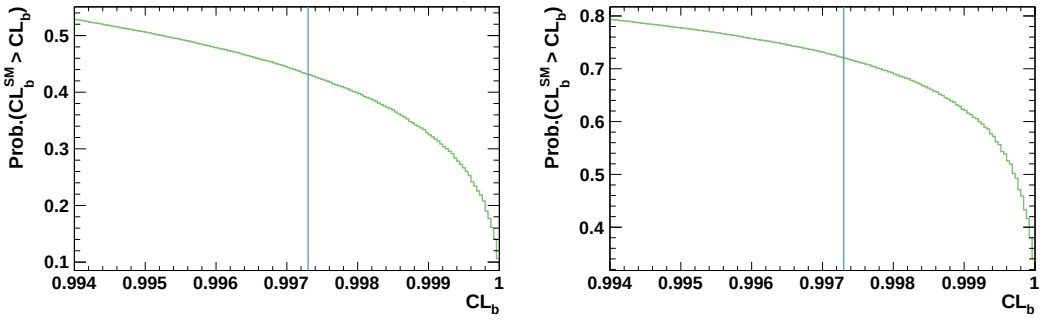
The $B^0 \to \mu^+\mu^-$ SM signal is expected to be compatible with the background only hypothesis at 38% in 2012, at 39% in 2011 and at 34% when the datasets are combined. Hence no SM signal hint is expected to be seen.

On the contrary, $B^0_s \to \mu^+\mu^-$ SM signal is expected to be compatible with the background only hypothesis at 0.5% in 2012, 1.1% in 2011 and 2×10^{-4} when both datasets are combined. Figure 3.36 shows the probability to excess a given CL_b value if the branching fraction is SM. It indicates that a SM signal evidence with a significance larger than 3 σ has a 72% (43%) probability to be observed in the combined (2012 only) dataset.

The sensitivity improvement between the 2012 and the 2011 analyses is due to a lower combinatorial background in the high BDT region and in

Table 3.9: Expected combinatorial background, $B^0_{(s)} \to h^+h'^-$ background, cross-feed, and signal events assuming SM predictions, together with the number of observed events in the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ mass signal regions, in bins of BDT for the 2011 data sample.

Mode	BDT bin	0.0 – 0.25	0.25 – 0.4	0.4 – 0.5	0.5 – 0.6	0.6 – 0.7	0.7 – 0.8	0.8 – 0.9	0.9 – 1
$B^0_s \to \mu^+\mu^-$	Exp. comb. bkg	1880^{+33}_{-33}	$55.5^{+3.0}_{-2.9}$	$12.1^{+1.4}_{-1.3}$	$4.16^{+0.88}_{-0.79}$	$1.81^{+0.62}_{-0.51}$	$0.77^{+0.52}_{-0.38}$	$0.47^{+0.48}_{-0.36}$	$0.24^{+0.44}_{-0.20}$
	Exp. peak. bkg	$0.129^{+0.066}_{-0.050}$	$0.066^{+0.024}_{-0.019}$	$0.052^{+0.018}_{-0.015}$	$0.047^{+0.015}_{-0.013}$	$0.053^{+0.017}_{-0.014}$	$0.050^{+0.016}_{-0.013}$	$0.052^{+0.017}_{-0.014}$	$0.049^{+0.018}_{-0.014}$
	Exp. signal	$2.70^{+0.81}_{-0.80}$	$1.30^{+0.27}_{-0.23}$	$1.03^{+0.20}_{-0.17}$	$0.92^{+0.15}_{-0.13}$	$1.06^{+0.17}_{-0.15}$	$1.10^{+0.17}_{-0.15}$	$1.26^{+0.20}_{-0.17}$	$1.31^{+0.28}_{-0.25}$
	Observation	1818	39	12	6	1	2	1	1
$B^0 \to \mu^+\mu^-$	Exp. comb. bkg	1995^{+34}_{-34}	$59.2^{+3.3}_{-3.2}$	$12.6^{+1.6}_{-1.5}$	$4.44^{+0.99}_{-0.86}$	$1.67^{+0.66}_{-0.54}$	$0.75^{+0.58}_{-0.40}$	$0.44^{+0.57}_{-0.38}$	$0.22^{+0.48}_{-0.20}$
	Exp. peak. bkg	$0.78^{+0.38}_{-0.29}$	$0.40^{+0.14}_{-0.10}$	$0.311^{+0.107}_{-0.079}$	$0.280^{+0.092}_{-0.068}$	$0.314^{+0.103}_{-0.076}$	$0.297^{+0.096}_{-0.071}$	$0.309^{+0.101}_{-0.075}$	$0.296^{+0.107}_{-0.079}$
	Exp. cross-feed	$0.43^{+0.13}_{-0.13}$	$0.205^{+0.044}_{-0.037}$	$0.163^{+0.032}_{-0.027}$	$0.145^{+0.025}_{-0.021}$	$0.168^{+0.029}_{-0.025}$	$0.174^{+0.029}_{-0.024}$	$0.199^{+0.033}_{-0.029}$	$0.206^{+0.046}_{-0.040}$
	Exp. signal	$0.328^{+0.096}_{-0.097}$	$0.158^{+0.030}_{-0.027}$	$0.125^{+0.022}_{-0.019}$	$0.112^{+0.016}_{-0.015}$	$0.129^{+0.019}_{-0.017}$	$0.134^{+0.018}_{-0.016}$	$0.153^{+0.022}_{-0.019}$	$0.159^{+0.032}_{-0.029}$
	Observation	1904	50	20	5	2	1	4	1



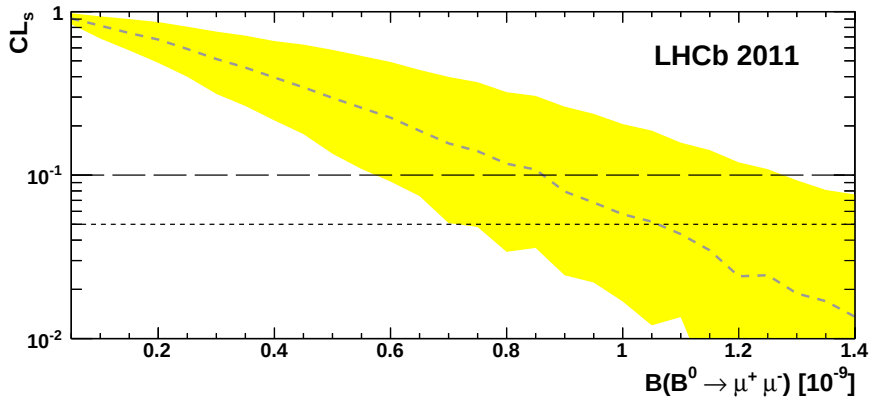
(a) (b)

Figure 3.36: Probability to have a CL_b larger than a given value if the signal was observed at the SM rate for the 2012 dataset (a) and combining with one from 2011 (b). The vertical blue lines indicates the CL_b value corresponding to a signal evidence with a significance of 3 σ

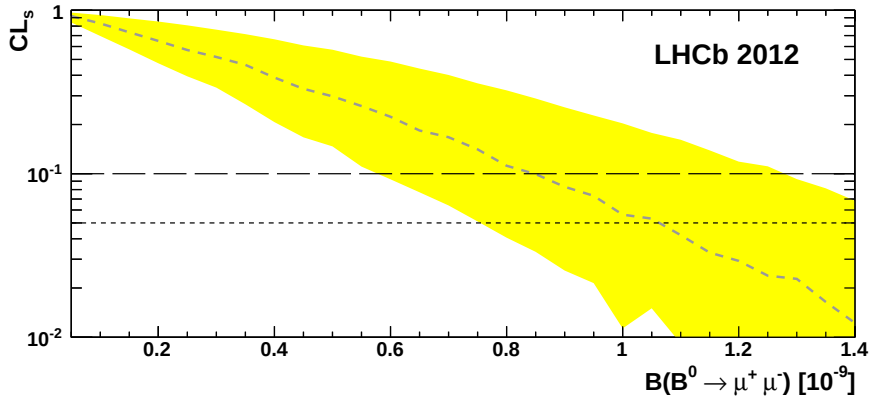
particular in the low mass sideband. Table 3.8 and Table 3.9 report indeed that in 2011 around 0.71 combinatorial background events are expected in the BDT region defined by $0.8 < \text{BDT} < 1$ whereas only 0.39 are expected in 2012 while the same amount of signal is expected in both analyses.

If no signal evidence is observed, an upper limit can be set using CL_s. For $B^0 \to \mu^+\mu^-$, Figure 3.37 shows the CL_s as a function of the branching fraction. If the signal is produced at the rate predicted by the SM, an upper limit on the branching fraction could be set at $9.6\ (7.6) \times 10^{-10}$ at 95 (90)% C.L. with the 2012 data, at $9.4\ (7.5) \times 10^{-10}$ at 95 (90)% C.L. with the 2011 data and at $6.0\ (5.0) \times 10^{-10}$ at 95 (90)% C.L. by combining the two.

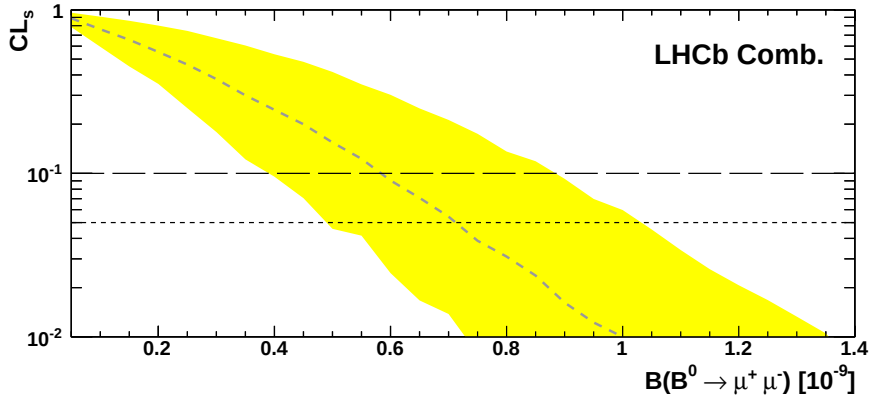
For $B_s^0 \to \mu^+\mu^-$ Figure 3.38 shows the CL_s as a function of the branching fraction hypothesis under the assumption of the background only and background plus SM signal. The figure shows that the CL_s distributions expected for each hypothesis are clearly distinct. The medians of these distributions are separated by more than two standard deviations for the 2011 analysis, which indicates that the analysis starts being able to distinguish a SM-like signal from a background like one. The separation is even larger for the 2012 analysis and overpasses three standard deviations when both analyses are combined. The worst case scenario would be that an SM like signal is observed but the pattern obtained belongs to the 28% for which the p-value is too large to claim an evidence. In this case, a combined upper limit around $6.3\ (5.6) \times 10^{-9}$ at 95 (90)% C.L. could be set. Another scenario corresponding to no signal evidence would be that the branching fraction of $B_s^0 \to \mu^+\mu^-$ is sensibly suppressed by some beyond standard model processes. In this case



(a)



(b)



(c)

Figure 3.37: CL_s as a function of the $B^0 \to \mu^+\mu^-$ branching fraction hypothesis. The gray dashed line inlaid in the yellow band is the median of the CL_s distributions obtained for each branching fraction hypothesis under the assumption of the SM signal and the band covers 34% on these distributions on each side of their median. Branching fraction hypotheses corresponding to CL_s smaller than the threshold represented by the horizontal dashed (dotted) line could be excluded at 90 (95)% C.L.

a combined upper limits could be set around 2.0 (1.6)$\times 10^{-9}$ at 95 (90)% C.L. excluding therefore the SM prediction.

In conclusion, we know that, even before looking at the data pattern, the analysis has reached a sensitivity which foresees potentially very interesting results. On one hand, if the signal strength is at least as large as the SM prediction, a signal evidence has a 72% probability to be observed. On the other hand, if the SM contributions to the decay interfere destructively with the one coming from NP, the SM prediction could be ruled out at more than 95% C.L.

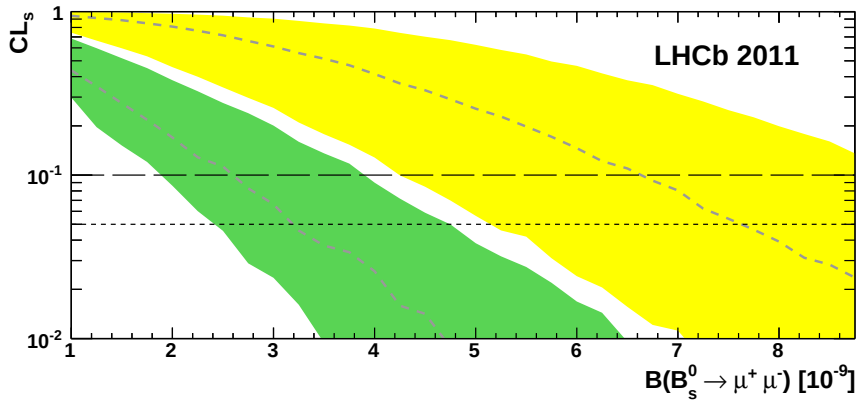
3.4.2 Observations

The data patterns observed in the 2011 and 2012 analyses are shown in Figure 3.39. In the two high BDT regions, some combinatorial and misidentified candidates ($B^0_{(s)} \to h^+h'^-$ and other exclusive modes) are expected to fall in the lower mass sideband. Then, according to our model, as the di-muon invariant mass increases, the number of combinatorial candidates decreases exponentially. Hence few of them are expected to be seen in the two signal mass region together with, for the B^0 mass window, some $B^0_{(s)} \to h^+h'^-$ misidentified candidates. Finally the exponential fall of the number of combinatorial candidates pursues in the upper mass sideband. Looking now at the observed data pattern, the sideband agrees with this smooth exponential fall but the sharp drop right after the B^0_s mass region hint at an additional non negligible signal contribution in the B^0_s mass region. This discussion can be more rigorously held by comparing the expected and observed BDT bin contents reported in Table 3.8 and Table 3.9. Strikingly, the 2012 $B^0_s \to \mu^+\mu^-$ last BDT bin contains 3 candidates while only $0.49^{+0.33}_{-0.21}$ backgrounds events are expected.

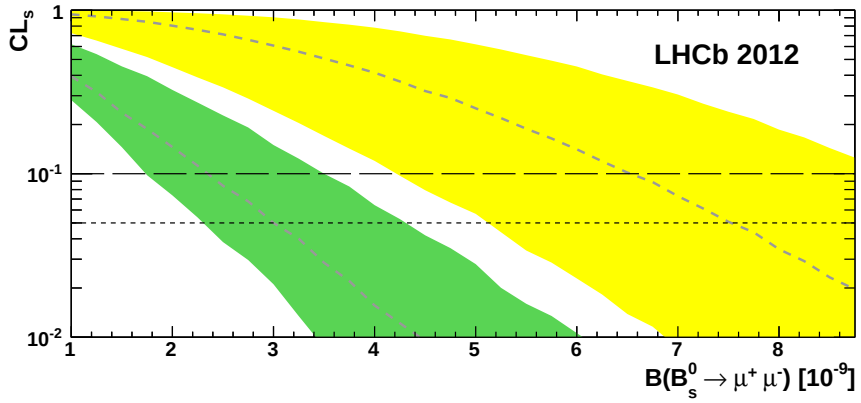
The CL_b allows to push further the comparison between the expected backgrounds and the observed pattern. For $B^0 \to \mu^+\mu^-$ the combined p-value is 11% and therefore too large to claim any signal hints. On the contrary, for $B^0_s \to \mu^+\mu^-$ already in the 2011 data the p-value is 9%. In the 2012 data, due to a significantly smaller background expectation for the same signal and observed number of events, the p-value is 9×10^{-4} which is already sufficient to claim an evidence of signal with a significance of 3.2 σ. Combining both analyses the p-value reaches 6×10^{-4} which corresponds to an evidence with a significance of 3.5 σ.

An estimate of the $B^0_s \to \mu^+\mu^-$ branching fraction can be obtained with the CL_s method by looking at the CL_{s+b}[8] curves shown in Figure 3.40. With

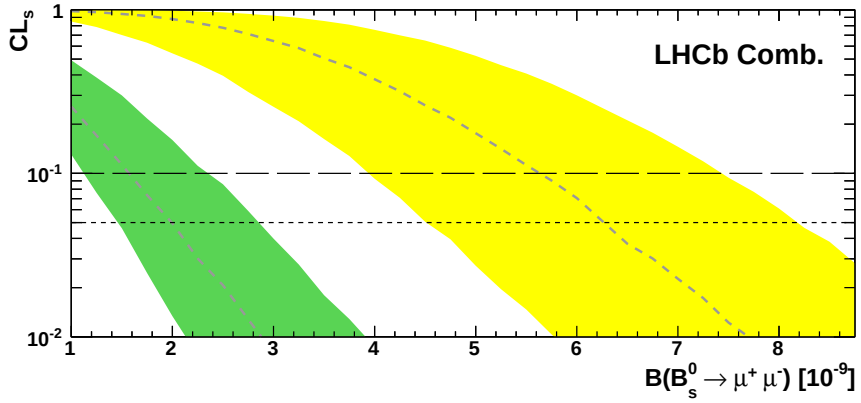
[8]Considering the small p-value value, the CL_s is almost equal to the CL_{s+b}.



(a)



(b)



(c)

Figure 3.38: CL_s as a function of the $B_s^0 \to \mu^+\mu^-$ branching fraction hypothesis. The gray dashed line inlaid in the yellow (green) band is the median of the CL_s distributions obtained for each branching fraction hypothesis under the assumption of the SM signal (backgound only) and the band covers 34% on these distributions on each side of their median. Branching fraction hypotheses corresponding to CL_s smaller than the threshold represented by the horizontal dashed (dotted) line could be excluded at 90 (95)% C.L.

(a)

(b)

Figure 3.39: Dimuon invariant mass versus BDT for selected candidates in the 2011 (a) and 2012 data (b); orange short-dashed (green long-dashed) lines indicate the $\pm 60\,\mathrm{MeV}/c^2$ search window around the B^0_s (B^0) mass.

the combined analyses, the branching fraction central value, obtained for $CL_{s+b} = 0.5$, is found to be 3.25×10^{-9} and the 95% C.L. interval, obtained for $CL_{s+b} = 0.025$ and $CL_{s+b} = 0.975$, reads $[1.05; 6.4]$. It is noteworthy to notice how well the observed CL_{s+b} curves agree with the SM expectations[9].

For $B^0 \to \mu^+\mu^-$, since no evidence of signal is observed, upper limits on the branching fraction can be set with the observed CL_s shown in Figure 3.41. The figures show good agreement between the expected and observed CL_s and the combined upper limit on the branching fraction is:

$$\mathcal{B}(B^0 \to \mu^+\mu^-) \;<\; 9.4\ (8.0) \times 10^{-10} \text{ at } 95\ (90)\% \text{ C.L.}, \qquad (3.15)$$

which is the world best upper limit on this decay.

3.4.3 Extraction of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$

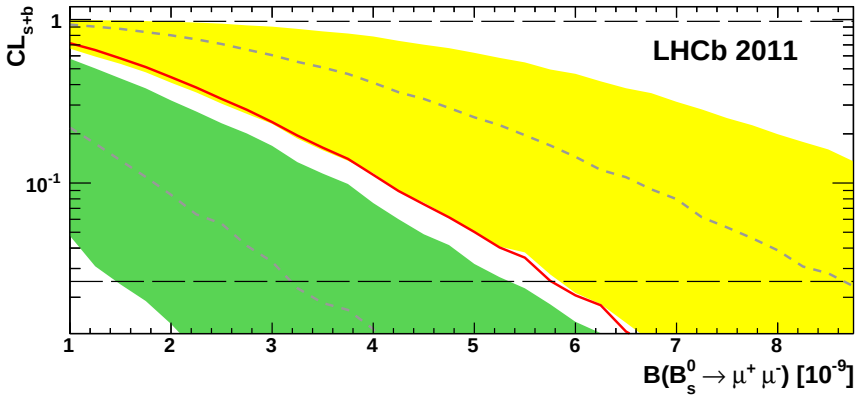
The results of the CL_s method indicate the presence of $B_s^0 \to \mu^+\mu^-$ signal candidates in the data. In order to extract the $B_s^0 \to \mu^+\mu^-$ branching fraction from these data, an un-binned maximum likelihood fit to the invariant mass distribution in all 2011 and 2012 BDT bins is performed. This fit uses the same PDFs as the ones used to derive the background yields from the mass sidebands and presented in Section 3.3.2. The likelihood is maximised with respect to the two signal branching fractions, the combinatorial background slope and yield, and all the other nuisance parameters are marginalised.

The log-likelihood of the fit profiled with respect to $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ is shown on Figure 3.42 and reaches its minimum at 3.2×10^{-9}. The 68% C.L. interval for $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ is defined as the one for which the log-likelihood is smaller than its minimum value plus 0.5 and is found to be $[2.0, 4.6] \times 10^{-9}$. Repeating the same process but fixing all nuisance parameters to their central values, allows to derive a second interval, reflecting the uncertainty only due to statistics. Finally, an additional ± 0.16 systematic uncertainty, accounting for the change of the combinatorial parametrisation from a single to double exponential is added. The combined results are:

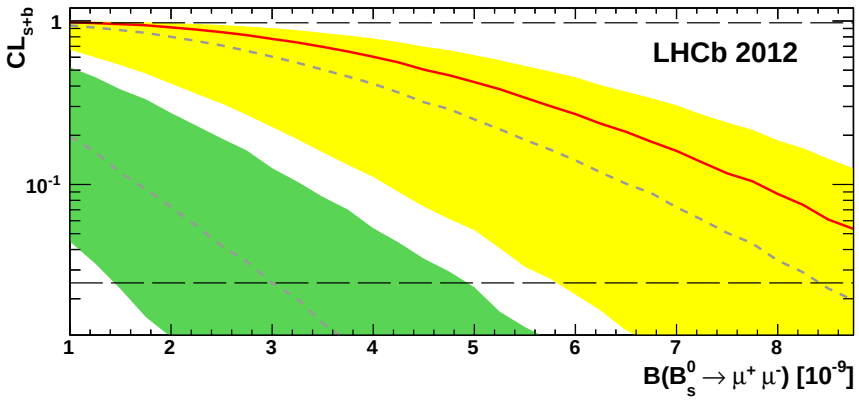
$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \;=\; (3.2^{+1.4}_{-1.2} \text{ stat. } {}^{+0.5}_{-0.3} \text{ syst.}) \times 10^{-9} \text{ at } 68\% \text{ C.L.} \qquad (3.16)$$

For a more detailed study of the systematic uncertainties, the reader is referred to [16]. This study indicates that, in the previous analysis [28], the dominant systematic uncertainty came from the assumption made on the background PDFs.

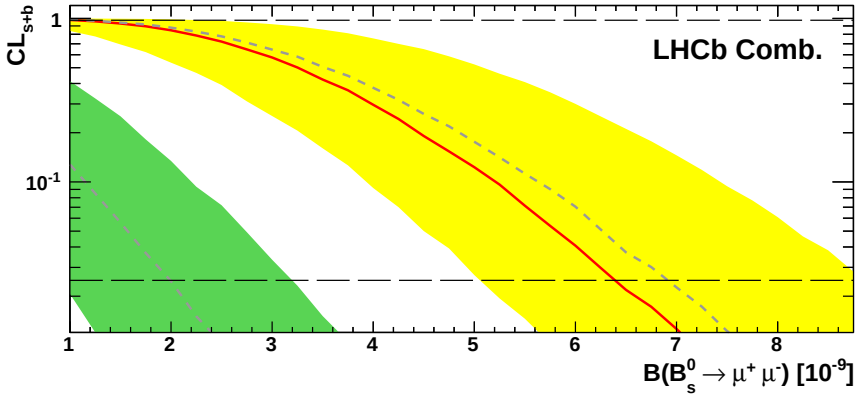
[9] As a cross check, we can notice that the expected CL_{s+b} is 0.5 when $\mathcal{B}(B_s^0 \to \mu^+\mu^-) \sim 3.6 \times 10^{-9}$ which is the SM prediction for the time integrated branching fraction.



(a)

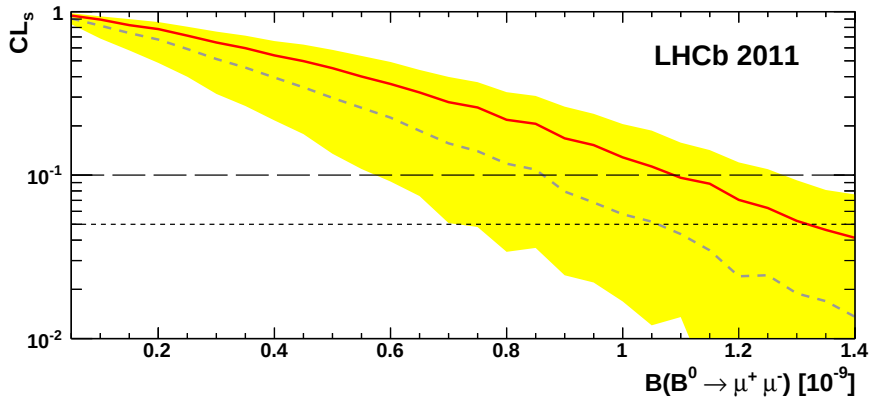


(b)

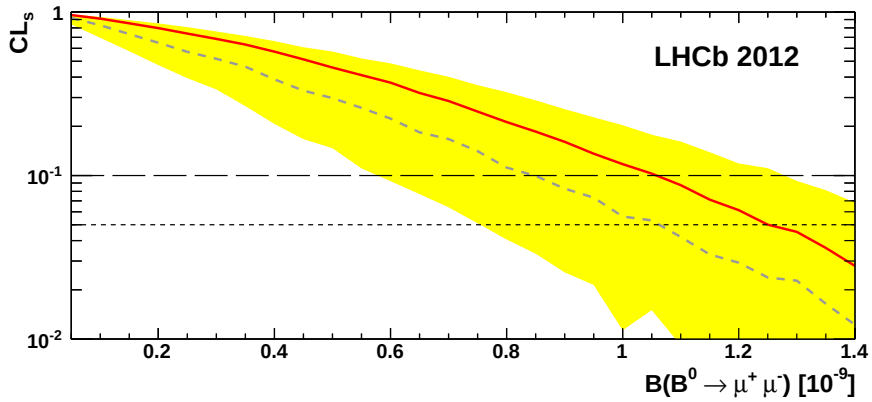


(c)

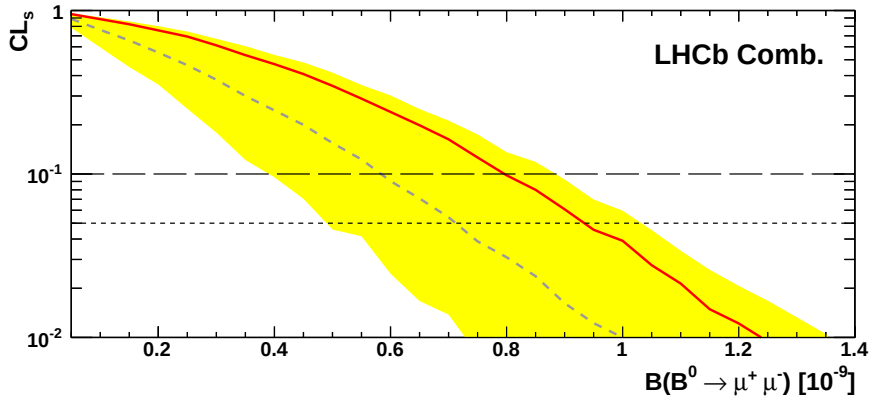
Figure 3.40: CL_{s+b} as a function of the $B_s^0 \to \mu^+\mu^-$ branching fraction hypothesis. The gray dashed line inlaid in the yellow (green) band is the median of the CL_{s+b} distributions obtained for each branching fraction hypothesis under the assumption of the SM signal (background only) and the band covers 34% on these distributions on each side of their median. The red line is the observed CL_{s+b} and its intersection with the two dashed horizontal lines defines an interval which contains at 95% C.L. the true branching fraction.



(a)



(b)



(c)

Figure 3.41: CL_s as a function of the $B^0 \to \mu^+\mu^-$ branching fraction hypothesis. The gray dashed line inlaid in the yellow band is the median of the CL_s distributions obtained for each branching fraction hypothesis under the assumption of the SM signal and the band covers 34% on these distributions on each side of their median. The red line is the observed CL_s. Branching fractions hypotheses corresponding to CL_s smaller than the threshold represented by the horizontal dashed (dotted) line are excluded at 90 (95)% C.L.

$B_s^0 \to \mu^+\mu^-$			2011	2012	Combined	Projection
p-value	SM+Bkg		1.1%	0.5%	2×10^{-4}	4×10^{-4}
	Observed		9%	9×10^{-4}	6×10^{-4}	
Median $\mathcal{B}$		10^{-9}	1.8	4.5	3.25	
Interval at 95% C.L.		10^{-9}	$[-; 5.75]$	$[1.15; 9.8]$	$[1.05; 6.4]$	
Limits at 95(90)% C.L.	Bkg Only	10^{-9}	3.2 (2.6)	3.0 (2.4)	2.0 (1.6)	2.0
	SM+Bkg	10^{-9}	7.6 (6.6)	7.5 (6.5)	6.3 (5.6)	6.6
	Observed	10^{-9}	5.1 (4.2)	8.9 (7.8)	5.8 (5.2)	

$B^0 \to \mu^+\mu^-$			2011	2012	Combined
p-value	SM+Bkg		39%	38%	34%
	Observed		16%	16%	11%
Limits at 95(90)% C.L.	Bkg Only	10^{-10}	9.4 (7.5)	9.6 (7.6)	6.0 (5.0)
	SM+Bkg	10^{-10}	10.5 (8.7)	10.5 (8.5)	7.1 (5.8)
	Observed	10^{-10}	13 (11)	12.5 (10.5)	9.4 (8.0)

Table 3.10: Summary table of the expectations and observed numbers obtained with the CL_s method.

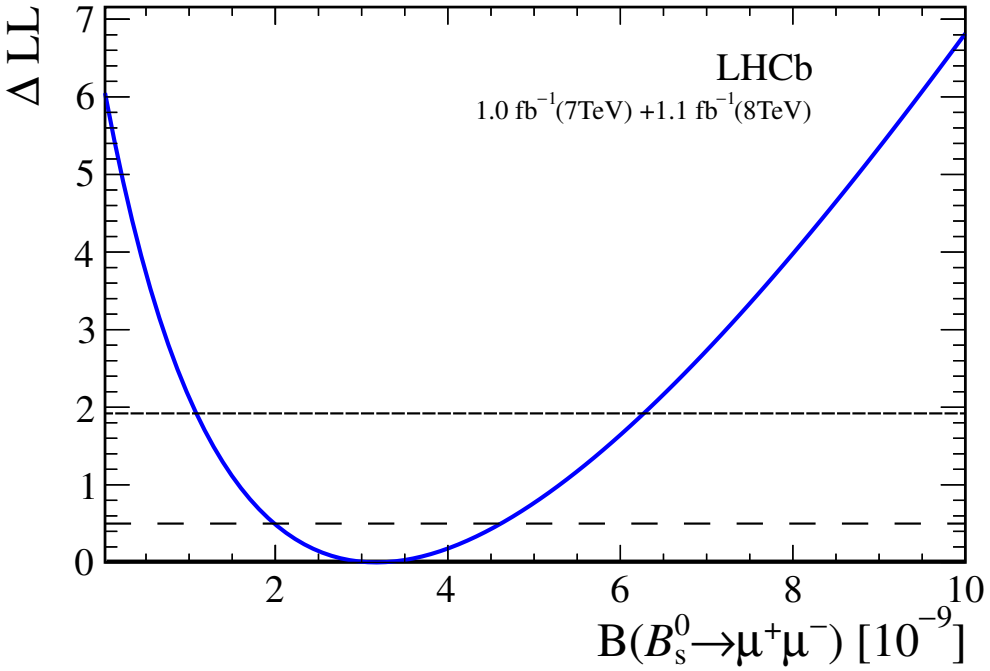


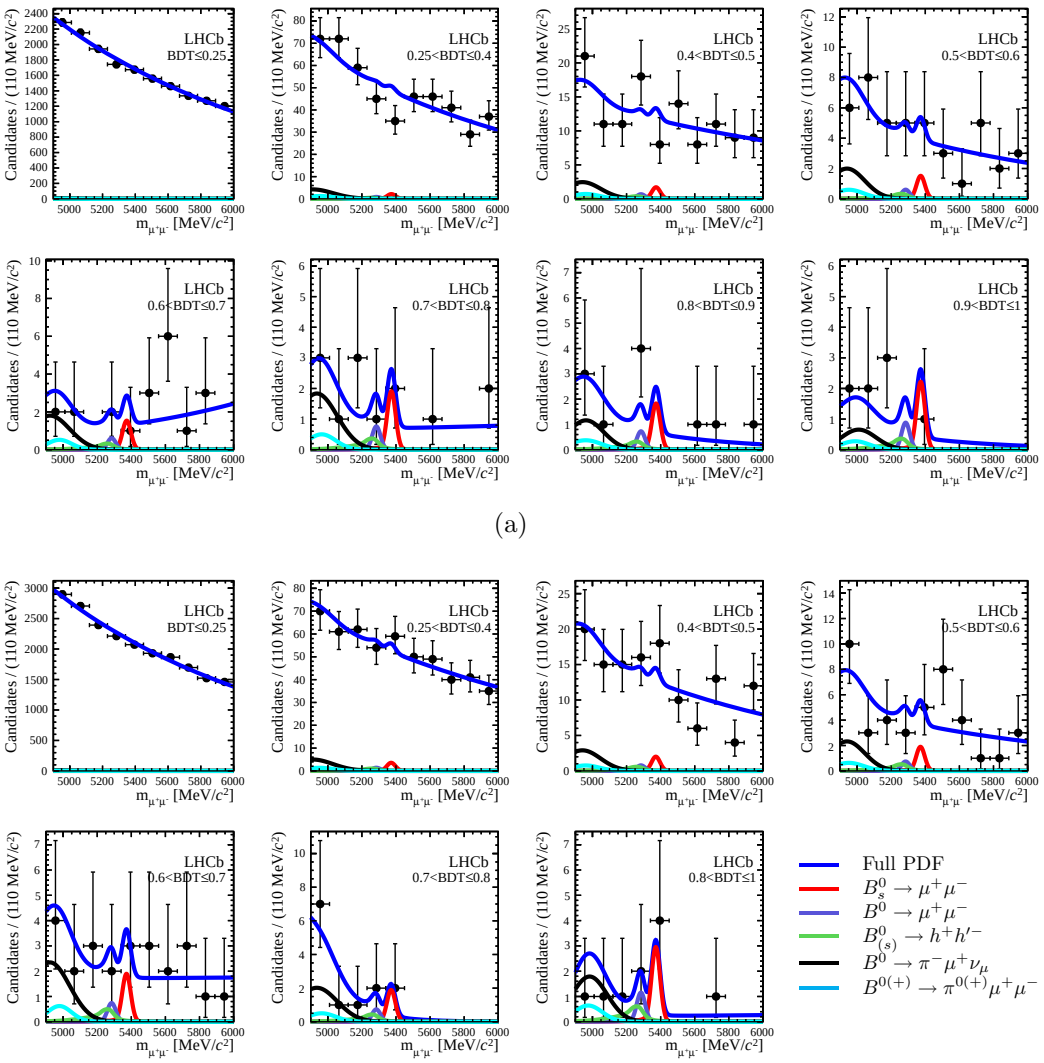
Figure 3.42: Log-likelihood of the fit of 2012 and 2011 data as a function of $\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-)$. Its intersection with the long (short)-dashed horizontal line defines an interval containing the true $\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-)$ at 68 (95) % C.L.

It is also worth noticing that the 95% C.L. that can be derived with the log-likelihood in Figure 3.42 is $[1.0; 6.3] \times 10^{-9}$ which agrees with the one derived with the $\mathrm{CL_s}$ method $[1.05; 6.4] \times 10^{-9}$.

The mass projection of the fit is shown in each 2011 and 2012 BDT bin in Figure 3.43 and Figure 3.44 shows the observed and fitted mass distribution for 2011 and 2012 data in the high BDT regions. In the BDT region larger than 0.5 or even 0.7, hints of signal can be inferred from the mass spectra. Going to the BDT region larger than 0.8, the signal hint is confirmed.

3.5 Combining with other LHC experiments

The general purpose LHC and Tevatron experiments are also contributing to the search for $\mathsf{B}^0_{(\mathsf{s})} \to \mu^+\mu^-$ decays. In particular, ATLAS and D0 looked for $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ [35, 36] and CMS and CDF for both $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ and $\mathsf{B}^0 \to$



(a)

(b)

Figure 3.43: Simultaneous fit of the invariant mass distribution in the 8 BDT bins of 2011 (a) and 7 BDT bins of 2012 (b) data.

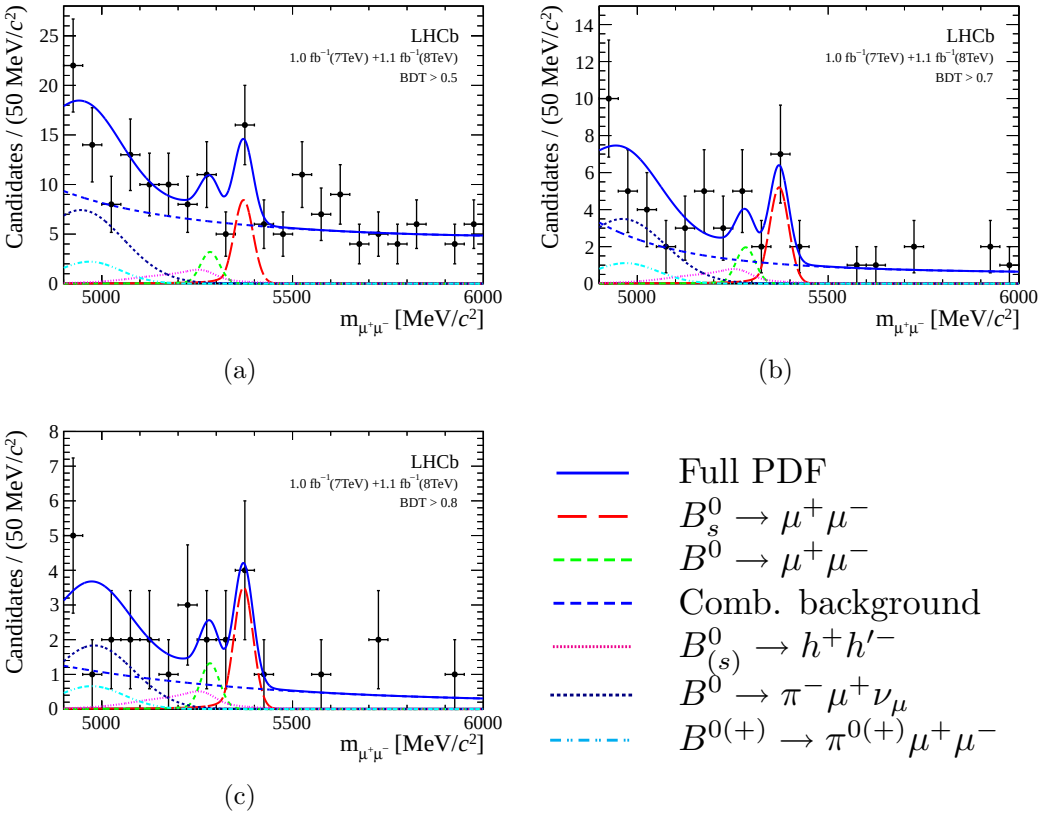


Figure 3.44: Invariant mass distributions of 2012 and 2011 candidates with a BDT larger than 0.5 (a), 0.7 (b), and 0.8 (c).

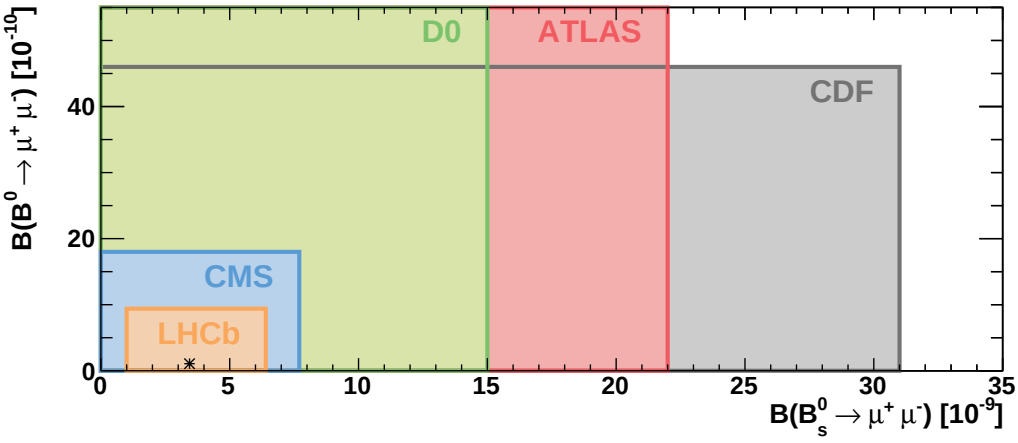


Figure 3.45: 95% C.L. experimental limits on the branching fraction of $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$. The star represents the SM central value prediction.

$\mu^+\mu^-$ [37, 38][10]. Except for CDF, which observed a small excess of event in the B_s^0 mass region with respect to the background expectation, all these experiments were only able to set upper limits on the branching fractions. All these results are compatible with each other and with those of LHCb and are reported in Figure 3.45.

All analyses are so far limited by statistics, therefore a substantial improvement is expected to be reached by combining them. The ATLAS, CMS, and the 2011 LHCb analyses have already been combined [42] and set upper limits on the branching fraction of $B_s^0 \to \mu^+\mu^-$ at 4.2×10^{-9} and at 8.1×10^{-10} for $B^0 \to \mu^+\mu^-$ which improved the best limits by 6% and 19% respectively. However these results, especially the $B^0 \to \mu^+\mu^-$ one, are known to be biased as they are derived from the original analysis version, published in [28], which did not account for the partially reconstructed background affecting the lower mass sideband which increased the background estimate in the signal region and consequently lowered the upper limits by 13% for $B_s^0 \to \mu^+\mu^-$ and 26% for $B^0 \to \mu^+\mu^-$. This section describes the update of this combination [43] which includes these corrections and the 2012 LHCb analysis.

The combination is performed with a method based on the modified frequentist approach, or CL_s method [24, 25]. The implementation of this method, already described in Section 3.1.9, is slightly different from the one developed by the CMS analysis based on a profile likelihood test statistic with

[10]The results presented at the 2013 EPS conference by ATLAS [39], CMS [40], and LHCb [41] are not taken into account here

profiled nuisance parameters (including B^0_s and B^0 crossfeed) and a sideband window sampling, and returns upper limits of $\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-) < 7.2 \times 10^{-9}$ and $\mathcal{B}(\mathsf{B}^0 \to \mu^+\mu^-) < 1.6 \times 10^{-9}$ at 95% C.L. instead of 7.7×10^{-9} and 1.8×10^{-10}.

All these analyses use $f_s/f_{d,u}$, the probability ratio that a b quark hadronises with an s quark to form a B^0_s or with a u or d quark to form a B^+ or a B^0, to normalise the $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ signal to control channels. However the values used are not the same in all analyses. The ATLAS, CMS, and LHCb 2011 analyses use $f_s/f_d = 0.267 \pm 0.021$ from [22][11] while the 2012 LHCb analysis uses a more up-to-date value, $f_s/f_d = 0.256 \pm 0.020$ [44]. In the combination we used this new value, hence all ATLAS, CMS, and LHCb 2011 $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ limits are expected to increase.

The ATLAS and CMS analyses are briefly described in Section 3.5.1 and Section 3.5.2 before presenting the results of the combination in Section 3.5.3.

3.5.1 ATLAS Analysis

The ATLAS $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ search, based on the first 2.4 fb^{-1} of 2011 data [35], is performed in three bins of $|\eta|_{\max}$, the maximum pseudorapidity ($|\eta|$) value of the two decay muons.

The likelihood used in the limit extraction with the CL$_\mathrm{s}$ method by ATLAS is given by:

$$\begin{aligned}\mathcal{L} \;=\; & G(\epsilon_{\mathrm{obs}}|\epsilon,\sigma_\epsilon)\, G(R^{\mathrm{b}}_{\mathrm{obs}}|R^{\mathrm{b}},\sigma_{R^{\mathrm{b}}}) \cdot \\ & \prod_{i=1}^{N_{\mathrm{bins}}} P(N^{\mathrm{s}}_{\mathrm{obs,i}}|\epsilon\,\epsilon_{\mathrm{i}}\,\mathcal{B} + N^{\mathrm{b}}_{\mathrm{i}} + N^{\mathrm{B}\to\mathrm{hh}}_{\mathrm{i}})\, P(N^{\mathrm{b}}_{\mathrm{obs,i}}|R^{\mathrm{b}}\,R^{\mathrm{b}}_{\mathrm{i}}\,N^{\mathrm{b}}_{\mathrm{i}})\, G(\epsilon_{\mathrm{obs,i}}|\epsilon_{\mathrm{i}},\sigma_{\epsilon,\mathrm{i}})\end{aligned} \tag{3.17}$$

with $N_{\mathrm{bins}} = 3$, $P(...)$ denoting the signal and background Poisson distributions and $G(...)$ the Gaussian distributions modelling the uncertainties of the nuisance parameters, i.e. of the global efficiency ϵ, of the background-to-signal-region ratio R and an additional Gaussian term describing the uncorrelated part of the errors on the relative efficiencies ϵ_i. N$^{\mathrm{b}}$ indicates the number of combinatorial background events, while $N^{\mathrm{B}\to\mathrm{hh}}$ indicates the number of peaking background events coming from misreconstructed $\mathsf{B}^0_{(\mathsf{s})} \to \mathsf{h}^+\mathsf{h}'^-$ events. The inverse of the product of the global (ϵ) and relative (ϵ_i) efficiencies[12] is the single-event-sensitivity (SES) per bin, i.e. SES $= (\epsilon\,\epsilon_{\mathrm{i}})^{-1}$.

Table 3.11 provides the ATLAS input values used in the CL$_\mathrm{s}$ extraction. It is important to note that the signal extraction is optimized using half

[11] ATLAS and CMS assume this value to be valid for their pseudorapidity range without adding any additional systematic.

[12] These efficiencies include signal detection efficiencies as well as normalization channel yields and cross-sections.

the background sample in the sidebands (events with odd event numbers) while the other half (events with even event numbers) is used to estimate the combinatorial background in the signal region from the sidebands

No ATLAS data is used in the determination of the combined limit on $\mathcal{B}(\mathsf{B}^0 \to \mu^+\mu^-)$.

3.5.2 CMS Analysis

Since the background level depends significantly on the pseudorapidity of the B candidate, the CMS analysis is performed separately in two channels, barrel and endcap, and then combined for the final result [37]. The barrel channel contains the candidates where both muons have $|\eta| < 1.4$ and the endcap channel contains those where at least one muon has $|\eta| > 1.4$. The expected number of combinatorial background events in the search window quoted in the table is extracted from a fit to the invariant mass sidebands. The contribution of the misidentified peaking background from $\mathsf{B}^0_{(s)} \to \mathsf{h}^+\mathsf{h}'^-$ in the B^0_s search window is very small, as it can be seen in Table 3.12. The only relevant correlation between uncertainties is due to the uncertainty on f_s/f_d, which is assumed to be 100% correlated between the number of expected signal events in the barrel and endcap regions. Other sources of correlation, such as the uncertainty on $\mathcal{B}(\mathsf{B}^+ \to \mathsf{J}/\psi\mathsf{K}^+)$, are negligible at the current level of precision.

3.5.3 Results and Conclusions

The p-value of the data observed in the $\mathsf{B}^0_s \to \mu^+\mu^-$ mass window for the background only hypothesis[13] is $(3.0 \pm 0.5) \times 10^{-4}$ which corresponds to 3.6σ. Figure 3.46 shows the $\mathrm{CL_{s+b}}$ as a function of the branching fraction and indicates that, if the observed data contain indeed some signal events, the true branching fraction belongs at 95% C.L. in this interval, $[1.1, 5.5] \times 10^{-9}$, and its best estimate ($\mathrm{CL_{s+b}} = 0.5$) is 3.0×10^{-9}. The analysis sensitivity can graphically be viewed as the separation between the background only and SM plus background expected $\mathrm{CL_{s+b}}$ versus branching fraction curves in Figure 3.46. Table 3.13 reports the p-values, branching fraction central values, 95% C.L. intervals, and upper limits set by each experiment and their combinations. For the ATLAS and CMS analyses the results obtained are slightly different from the ones published. These differences come from the

[13]The p-value is evaluated at the branching fraction hypothesis for which the $\mathrm{CL_{s+b}}$ is 0.5. Around 125k pseudo-experiments have been run.

Table 3.11: Input values used for the extraction of the $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ upper limit using the $\mathrm{CL_s}$ method from the ATLAS analysis of the first 2.4 fb^{-1} of the 2011 data [35]. The values are given for three bins in $|\eta|_{\max}$, the maximum $|\eta|$ value of the two signal decay muons: barrel ($|\eta|_{\max} < 1.0$), transition ($1.0 < |\eta|_{\max} < 1.5$) and endcap ($1.5 < |\eta|_{\max} < 2.5$) region. The signal region is defined to be the mass ranges 5250 – 5482, 5234 – 5498 and 5195 – 5537 MeV/c^2 for the three bins, respectively. The quoted uncertainties include both statistical and systematic sources, combined in quadrature. Parameters indicated as constants (const.) have no errors, either ($R_\mathrm{i}^\mathrm{bkg}$) because the uncertainty is common to the three bins and included in the corresponding global term, or ($N^{\mathrm{B}\to\mathrm{hh}}$) because the uncertainty does not affect the final result.

Region	Barrel	Transition	Endcap
Global efficiency ϵ	$(4.45 \pm 0.45) \cdot 10^3$		
Global bkg. scaling factor R^bkg	1.00 ± 0.04		
Relative efficiencies ϵ_i $[10^4]$	3.14 ± 0.17	1.40 ± 0.15	1.58 ± 0.26
Relative bkg. scaling factors $R_\mathrm{i}^\mathrm{bkg}$ (const.)	1.29	1.14	0.88
Observed in signal region $N_\mathrm{i}^\mathrm{sig}$	2	1	0
Observed in sideband regions $N_\mathrm{i}^\mathrm{bkg}$	5	0	2
Exp. peaking bkg. in sig. region $N_\mathrm{i}^{\mathrm{B}\to\mathrm{hh}}$ (const.)	0.10	0.06	0.08

Table 3.12: Expected background events (excluding misidentification), expected background events from misidentification, expected signal events assuming the SM branching fraction prediction, and observed events in the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ search windows, from the CMS analysis of the 2011 data. The mass range 5200 – 5300 MeV/c^2 is associated with the B^0 window and 5300 – 5450 MeV/c^2 with the B_s^0 window. Uncertainties include systematic effects.

Invariant Mass (MeV/c^2)		Barrel region	Endcap region
5200 – 5300	Exp. bkg.	0.40 ± 0.34	0.76 ± 0.35
	Exp. misid.	0.33 ± 0.07	0.15 ± 0.03
	Exp. signal	0.24 ± 0.02	0.10 ± 0.01
	Observed	2	0
5300 – 5450	Exp. bkg.	0.59 ± 0.50	1.14 ± 0.53
	Exp. misid.	0.18 ± 0.06	0.08 ± 0.02
	Exp. signal	2.70 ± 0.41	1.23 ± 0.18
	Observed	2	4

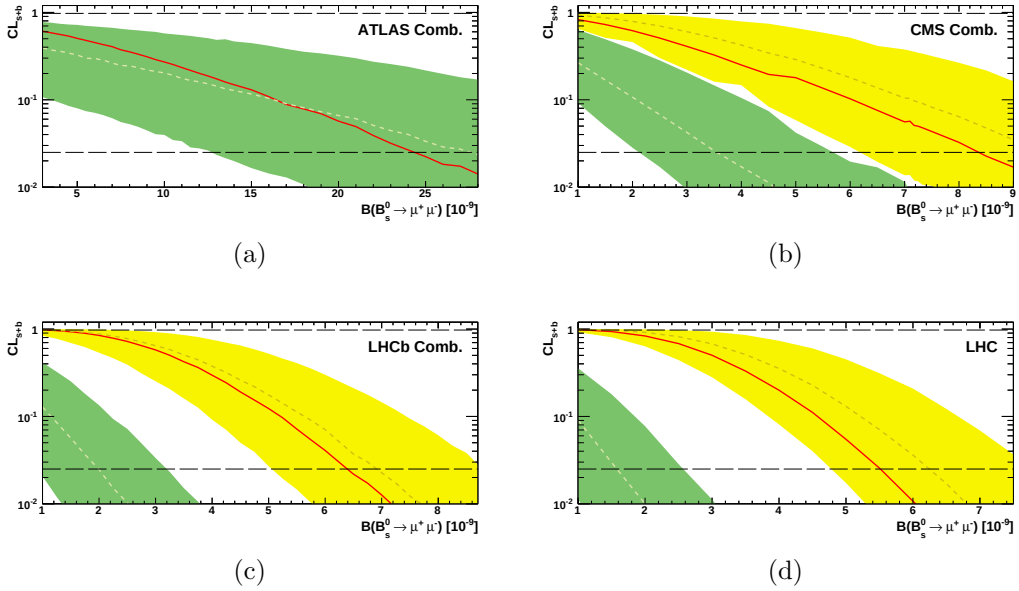
update of f_s/f_d (4%), of the SM branching fraction prediction (9%), and from different implementations of the CL_s method.

The p-value of the data observed in the $B^0 \to \mu^+\mu^-$ mass window for the background only hypothesis is $(7.3 \pm 0.1)\%$. Hence an upper limit is set at

$$\mathcal{B}(B^0 \to \mu^+\mu^-) \;<\; 8.1 \times 10^{-10}, \tag{3.18}$$

with the CL_s which is shown as a function of the branching fraction in Figure 3.47.

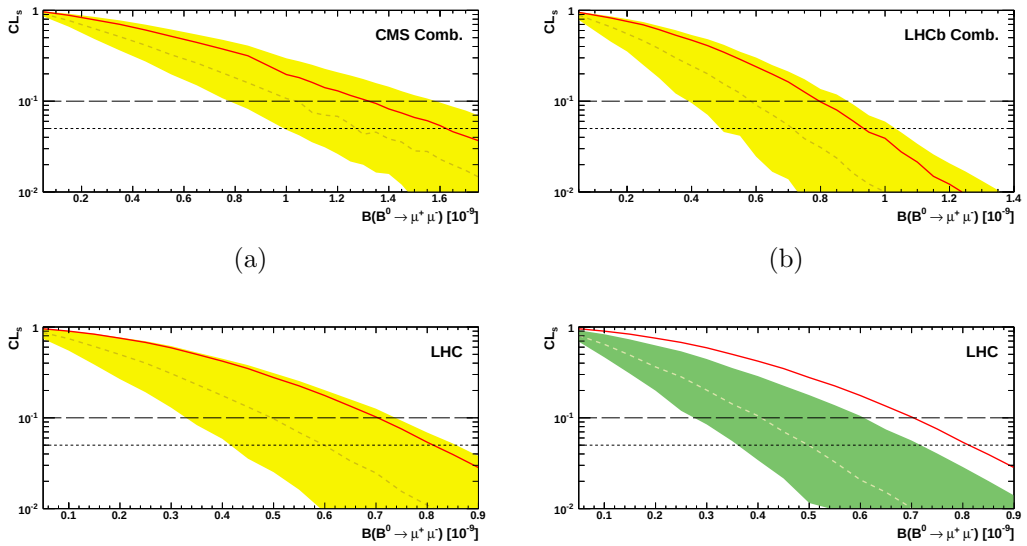
This results improves the constraints brought by any individual experiment reported in Table 3.14 by at least 14%, and is the world best upper limit. Like for $B_s^0 \to \mu^+\mu^-$, the results obtained for ATLAS and CMS are slightly different from the one published, due to different implementations of the CL_s method. The previous combination [42], reported the same limits but the partially reconstructed background candidates affecting the lower mass sideband in the 2011 LHCb analysis were ignored which increased the expected combinatorial background in the signal region and lowered artificially the limits by 26%. The gain in sensitivity obtained by combining can be measured by the improvement of the expected limits under the assumption of the SM plus background. This limit is $\mathcal{B}(B^0 \to \mu^+\mu^-) < 6.0 \times 10^{-10}$ at 95% C.L. which is 15% lower than the combined limits brought by the 2011 and 2012 LHCb analyses.



(a) (b)

(c) (d)

Figure 3.46: CL_{s+b} as a function of the $B_s^0 \to \mu^+\mu^-$ branching fraction for ATLAS (a), CMS (b), LHCb 2011 and 2012 (c), and the combination (d). The dashed line inlaid in the yellow (green) band is the median of the CL_{s+b} distributions obtained for each branching fraction hypothesis under the assumption of the SM signal (background only) and the band covers 34% of these distributions on each side of their median. The red line is the observed CL_{s+b} and its intersection with the two dashed horizontal lines defines an interval containing at 95% C.L. the true branching fraction.



(a) (b)

(c) (d)

Figure 3.47: CL_s as a function of the $B^0 \to \mu^+\mu^-$ branching fraction for CMS (a), LHCb 2011 and 2012 (b), and the combination (c and d). The gray dashed line inlaid in the yellow (green) band is the median of the CL_s distributions obtained for each branching fraction hypothesis under the assumption of the SM plus background (background only) and the band covers 34% of these distributions on each side of their median. The red line is the observed CL_s. Branching fractions corresponding to a CL_s smaller than the threshold represented by the horizontal dashed (dotted) line are excluded at 90 (95)% C.L.

Table 3.13: Summary table of the expectations and observed results obtained with the CL_s method for $B_s^0 \to \mu^+\mu^-$.

	p-value		$\mathcal{B}$ Median 10^{-9}	$\mathcal{B}$ Interval 95 % C.L. 10^{-9}	$\mathcal{B}$ Upper Limits 95 % C.L. 10^{-9}		
	SM+Bkg	Obs	Obs	Obs	SM+Bkg	Bkg Only	Obs
ATLAS	35%	30%	–	–	–	26	24
ATLAS Published	–	44%	–	–	–	23	22
CMS	1.1%	3.7%	2.5	[–; 8.4]	8.4	3.6	7.3
CMS Published	–	11%	–	–	8.4	–	7.7
LHCb 2011	1.1%	9.0%	1.8	[–; 5.7]	7.6	3.2	5.1
LHCb 2012	5.1×10^{-3}	8.6×10^{-4}	4.5	[1.1; 9.8]	7.5	3.0	8.9
LHCb Combined	2.4×10^{-4}	6.1×10^{-4}	3.3	[1.0; 6.4]	6.3	2.0	5.8
Combined	4.8×10^{-5}	3.0×10^{-4}	3.0	[1.1; 5.5]	5.8	1.6	5.0

Table 3.14: Summary table of the expectations and observed results obtained with the CL_s method for $B^0 \to \mu^+\mu^-$.

	p-value		$\mathcal{B}$ Upper Limits 95 % C.L. 10^{-10}		
	SM+Bkg	Obs	SM+Bkg	Bkg Only	Obs
CMS	33%	13%	13	13	16
CMS Published	–	24%	16	–	18
LHCb 2011	38%	19%	11	9.4	13
LHCb 2012	38%	16%	11	9.6	13
LHCb Combined	34%	11%	7.1	6.1	9.4
Combined	31%	7.3%	6.0	5.0	8.1

3.6 Conclusion

This chapter described the analysis of data corresponding to the $2.1\,\text{fb}^{-1}$ of integrated luminosity recorded by the LHCb detector in 2011 and 2012 [19]. Four main steps can be identified in the analysis strategy. First a loose selection is performed to remove obvious background events. The selected events are subsequently classified in a plane made of the candidate invariant mass and of a multivariate classifier, called BDT, based on kinematic and topological characteristics of the candidates. The signal and background expected populations in this plane are derived with data driven techniques. Finally, the compatibility of these expectations with the observed data are derived with the CL_s method.

The analysis sensitivity is determined in the two first steps of the analysis (event selection and classification) and in particular the second one. Once the variables used to separate signal and background events are defined, the choice of the categories in which events are classified is crucial. An original method was developed to choose these categories in order to optimise the analysis sensitivity. The optimal binning improved the analysis [26] sensitivity by 14%.

This method was designed bearing in mind the CL_s method and its main figure of merit was built from the test statistic. The CL_s method in itself was briefly introduced at the beginning of the section. Its implementation for the analysis was described subsequently to great length. Particular attention was given to the treatment of the statistical and systematic uncertainties. On the basis of the previous analysis [28], sensitivity projections were obtained.

These projections are in agreement with the sensitivity obtained in the new analysis [19]. Such a sensitivity allows to obtain the first evidence of the $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ decay with a significance of 3.5 σ and therefore to measure it branching fraction. This measurement,

$$\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-) \;=\; (3.2^{+1.4}_{-1.2}\ \text{stat.}\ ^{+0.5}_{-0.3}\ \text{syst.}) \times 10^{-9}, \tag{3.19}$$

is in agreement with the SM prediction. For $\mathsf{B}^0 \to \mu^+\mu^-$, no compelling signal hint is found and the decay branching fraction is constrained to be lower than 9.4 (8.0) $\times\, 10^{-10}$ at 95 (90)% C.L..

These results follow and supersede previous searches performed with the ATLAS, CMS, and LHCb experiments. However these past results remain relevant and larger $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ signal significance and lower $\mathsf{B}^0 \to \mu^+\mu^-$ branching ratio upper limit can be obtained by combining the new LHCb results with those of ATLAS and CMS. This combination has been performed and brings the $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ significance at 3.6σ. Again, no compelling $\mathsf{B}^0 \to$

$\mu^+\mu^-$ signal hint is observed, and the upper limit on the decay branching fraction is lowered at 8.1×10^{-10} at 95 (90)% C.L.

The first evidence of $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ is a milestone in the worldwide and nineteen years long effort to search for $\mathsf{B}^0_{(\mathsf{s})} \to \mu^+\mu^-$ decays. The branching fraction measurement is in agreement with the SM prediction. Despite the large experimental uncertainties, this measurement indicates that searches of new physics in these modes require precise measurements in which all tiny and subtle effects must be considered. In Section 1, we discussed the phenomenology related to the B-mixing and introduced a new observable, $\mathcal{A}_{\Delta\Gamma}$. This observable is an essential parameter for the decay-time distribution of $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$. As long as it is not measured, its value must be postulated. As a consequence, the next section, will show that the analysis is model dependent and that accounting for this model dependency allows to gain some sensitivity to the search of physics beyond the SM.

Bibliography

[1] R. Giles et al. *Two-Body Decays of* B *Mesons.* Physical Review, **D30** (1984), 2279.

[2] LHCb Collaboration. `http://cern.ch/LHCb-release-area/DOC/gauss/`.

[3] T. Sjostrand, et al. *PYTHIA 6.4 Physics and Manual.* Journal of High Energy Physics, **0605** (2006), 026.

[4] T. Sjostrand, et al. *A Brief Introduction to PYTHIA 8.1.* Computer Physics Communications, **178** (2008), 852–867.

[5] J. Lange and others. `http://www.slac.stanford.edu/~lange/EvtGen/`,.

[6] J. Allison et al. *Geant4 developments and applications.* IEEE Transactions on Nuclear Science, **53** (2006) 1, 270–278.

[7] LHCb Collaboration. `http://cern.ch/LHCb-release-area/DOC/boole/`.

[8] LHCb Collaboration. `http://cern.ch/LHCb-release-area/DOC/brunel/`.

[9] LHCb Collaboration. `http://cern.ch/LHCb-release-area/DOC/moore/`.

[10] S. Kullback et al. *On information and sufficiency.* The Annals of Mathematical Statistics, **22** (1951) 1, 79–86.

[11] S. Kullback. *The kullback-leibler distance.* The American Statistician, **41** (1987) 4, 340–341.

[12] M. Needham. *Clone Track Identification using the Kullback-Liebler Distance.* CERN-LHCb-2008-002 (2008).

[13] G. Lanfranchi, et al. *The Muon Identification Procedure of the LHCb Experiment for the First Data.* (2009) LHCb-PUB-2009-013. CERN-LHCb-PUB-2009-013.

[14] LHCb Collaboration. *Central Exclusive Dimuon Production at* $\sqrt{s} = 7$ *TeV.* LHCb-CONF-2011-002 (2011).

[15] Y. Aditya, et al. *Faking* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$. Physical Review, **D87** (2013), 074028.

[16] C. Adrover. *Towards the first* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *measurements with the LHCb detector*. Ph.D. thesis, Aix-Marseille U. (2012). CERN-THESIS-2012-123, Presented 10 Sep 2012.

[17] CDF Collaboration. *Search for* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *and* $\mathsf{B}^0 \to \mu^+\mu^-$ *decays in* $\mathsf{p\overline{p}}$ *collisions with CDF II*. Physical Review Letter, **95** (2005), 221805.

[18] T. Skwarnicki. *A study of the radiative cascade transitions between the upsilon-prime and upsilon resonances*. Ph.D. thesis, Cracow TU, Hamburg (1986).

[19] LHCb Collaboration. *First evidence for the decay* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$. Physical Review Letter, **110** (2013), 021801.

[20] B. Adeva et al. *Roadmap for selected key measurements of LHCb*. (2009).

[21] PDG. *Review of particle physics*. Journal of Physics, **G37** (2010), 075021.

[22] LHCb Collaboration. *Measurement of b-hadron production fractions in* 7 TeVpp *collisions*. Physical Review, **D85** (2012), 032008.

[23] J. A. H. Morata, et al. *Measurement of trigger efficiencies and biases*. CERN-LHCb-2008-073 (2010).

[24] T. Junk. *Confidence level computation for combining searches with small statistics*. Nuclear Instruments and Methods in Physics Research, **A434** (1999), 435–443.

[25] A. L. Read. *Presentation of search results: The CL_s technique*. Journal of Physics, **G28** (2002), 2693–2704.

[26] LHCb Collaboration. *Search for the rare decays* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *and* $\mathsf{B}^0 \to \mu^+\mu^-$. Physics Letter, **B708** (2012), 55–67.

[27] M. Perrin-Terrin et al. *Optimisation of the binning of the discriminating variables used in the computation of* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *upper limits with the modified frequentist approach*. LHCb-INT-2012-003 (2012).

[28] LHCb Collaboration. *Strong constraints on the rare decays* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *and* $\mathsf{B}^0 \to \mu^+\mu^-$. Physical Review Letter, **108** (2012), 231801.

[29] LHCb Collaboration. *Searches for violation of lepton flavour and baryon number in tau lepton decays at LHCb.* Physics Letter, **B724** (2013), 36–45.

[30] S. S. Wilks. *The large-sample distribution of the likelihood ratio for testing composite hypotheses.* The Annals of Mathematical Statistics, (1938), 60–62.

[31] W. Verkerke et al. *The RooFit toolkit for data modeling.* Electronic Conference Proceedings Archive, **C0303241** (2003), MOLT007.

[32] T. Junk et al. *Sensitivity, exclusion and discovery with small signals, large backgrounds, and large systematic uncertainties.* CDF Public note, (2007) 8128.

[33] L. Moneta, et al. *The RooStats Project.* ArXiv:1009.1003, (2010).

[34] J. R. Priest. *An Interesting Relation between the Average and Most Probable Value in a Poisson Probability Distribution.* American Journal of Physics, **38** (1970), 658–659.

[35] G. Aad et al. *Search for the decay* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *with the ATLAS detector.* Physics Letter, **B713** (2012), 387–407.

[36] Abazov et al. *Search for the rare decay* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$. Physical Review, **87** (2013) 7, 072006.

[37] CMS Collaboration. *Search for* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *and* $\mathsf{B}^0 \to \mu^+\mu^-$ *decays in pp collisions at 7 TeV.* Physical Review Letter, **107** (2011), 191802.

[38] CDF Collaboration. *Search for* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *and* $\mathsf{B}^0 \to \mu^+\mu^-$ *decays with the full CDF Run II data set.* Physical Review, **D87** (2013) 7, 072003.

[39] *Limit on* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *branching fraction based on* 4.9 fb^{-1} *of integrated luminosity.* ATLAS-CONF-2013-076 (2013).

[40] CMS Collaboration. *Measurement of the* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *branching fraction and search for* $\mathsf{B}^0 \to \mu^+\mu^-$ *with the CMS Experiment.* Physical Review Letter, **111** (2013), 101804.

[41] LHCb Collaboration. *Measurement of the* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *branching fraction and search for* $\mathsf{B}^0 \to \mu^+\mu^-$ *decays at the LHCb experiment.* Physical Review Letter, **111** (2013) 10, 101805.

[42] C. ATLAS et al. *Search for the rare decays* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *at the LHC with the ATLAS, CMS and LHCb experiments.* LHCb-CONF-2012-017 (2012).

[43] G. M. M. Perrin-Terrin. *Search for the rare decays* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *at the LHC with the ATLAS, CMS and LHCb experiments.* LHCb-INT-2013-020 (2013).

[44] LHCb Collaboration. *Measurement of the fragmentation fraction ratio* f_s/f_d *and its dependence on* B *meson kinematics.* Journal of High Energy Physics, **1304** (2013), 001.

Chapter 4

Interpretation of the Results

The results of the LHCb analysis [1] and its combination with the two other LHC experiments point close towards the SM prediction. Hence, as already stated in Section 1.4, the comparison between experimental results and predictions must be performed very carefully. For this purpose, Section 4.1 describes the effects due to the B-mixing and explains that the experimental results are model dependent. Section 4.2 starts by giving a review of some model independent comparisons between theoretical predictions and experimental results found in the literature. Then, these comparisons are refined to account for the model dependency established before. Finally the chapter ends with a brief model dependent discussion focused on super-symmetric models.

4.1 Model Dependent Results

The interplay between mixing and decay in the $B_s^0 \to \mu^+\mu^-$ channel is parametrised by $\mathcal{A}_{\Delta\Gamma}$ as explained in Section 1.3 and this observable can be affected by new physics. Thus, as long as $\mathcal{A}_{\Delta\Gamma}$ is not measured, a degree of freedom is left in the analysis. Section 4.1.1 explains the process through which the value of $\mathcal{A}_{\Delta\Gamma}$ affects the analysis. This process is acting in the normalisation to control channels, and in the calibration of the classifier PDF. These two effects are described in Section 4.1.2 and Section 4.1.3 for the normalisation and in Section 4.1.4 for the calibration. Finally, in Section 4.1.5, the experimental results are corrected to account for the $\mathcal{A}_{\Delta\Gamma}$ dependency. All studies presented in this section were initially documented in [2].

4.1.1 B_s^0-$\overline{B}_s^0$ Mixing Effect

The time dependent decay rate of $B_s^0 \to \mu^+\mu^-$ is the sum of two exponential functions accounting for the fact that the heavy and light B_s^0 mass eigenstates have different decay widths (see Equation 1.12). However, the relative fraction of the two exponential functions is not known as it depends on the model being considered. Thus this distribution is described by a model dependent factor $\mathcal{A}_{\Delta\Gamma}$ defined in Equation 1.15 and two parameters measured experimentally, $\tau_{B_s^0}$ and y_s, defined in Equation 1.14. In the SM, $\mathcal{A}_{\Delta\Gamma} = 1$ so only the heavy mass eigenstate decays to a di-muon final state. The decay time distribution is therefore a single exponential function of slope Γ_H.

Subtle effects arise in the analysis when selection criteria involving the candidate decay time are applied. In this case, the time integrated efficiency,

$$\epsilon = \frac{\int_0^\infty \Gamma(B_s^0(t) \to \mu^+\mu^-, \mathcal{A}_{\Delta\Gamma}, y_s)\epsilon(t)\mathrm{d}t}{\int_0^\infty \Gamma(B_s^0(t) \to \mu^+\mu^-, \mathcal{A}_{\Delta\Gamma}, y_s)\mathrm{d}t}, \tag{4.1}$$

becomes model dependent. However, note that in Equation 4.1, the time dependent efficiency, $\epsilon(t)$, is not model dependent. In practice, the time integrated efficiency is obtained from simulated data which necessarily assume some shape for the time dependent decay rate. Hence to translate the time integrated efficiency obtained with simulated data to another model a correction factor must be applied.

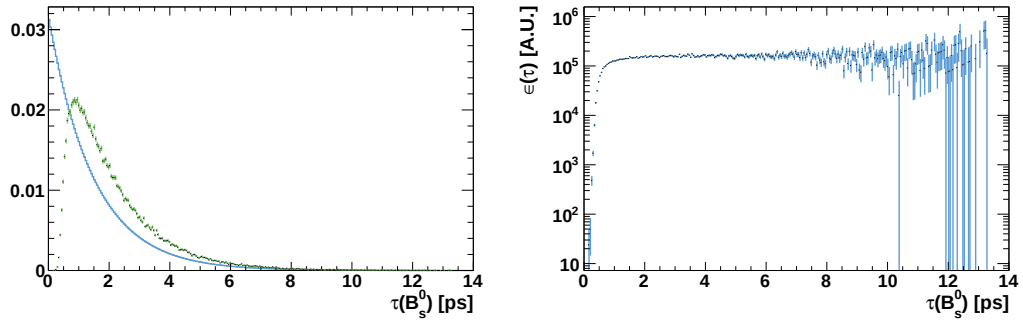
4.1.2 Model Dependent Normalisation

The first effect in the analysis arises when the time integrated efficiency of the initial selection described in Section 3.1.4 is computed. As shown in Table 3.2, the signal selection criteria involve variables highly correlated with the candidate decay time like the impact parameter of the candidate tracks. Hence the selection efficiency is decay time dependent as shown in Figure 4.1.

The signal simulated data time dependent decay rate used in the analysis, is a single exponential of slope $1/1.469141\,\mathrm{ps}^{-1}$. Hence, in order to translate the time integrated efficiency obtained with these data to another model (i.e. another $\mathcal{A}_{\Delta\Gamma}$), the following factor is defined[1]:

$$\begin{aligned}\delta_\epsilon &= \frac{\epsilon^{\mathcal{A}_{\Delta\Gamma}, y_s}}{\epsilon^{MC}} \\ &= \frac{\int_0^\infty \Gamma(B_s^0(t) \to \mu^+\mu^-, \mathcal{A}_{\Delta\Gamma}, y_s)\epsilon(t)\mathrm{d}t}{\int_0^\infty \Gamma(B_s^0(t) \to \mu^+\mu^-, \mathcal{A}_{\Delta\Gamma}, y_s)\mathrm{d}t} \times \frac{\int_0^\infty e^{-\Gamma_{MC}t}\mathrm{d}t}{\int_0^\infty e^{-\Gamma_{MC}t}\epsilon(t)\mathrm{d}t}.\end{aligned} \tag{4.2}$$

[1] In this expression we also consider the y_s dependency as the uncertainty on the experimental results are still large (10%) some significant shift can be expected.



(a) (b)

Figure 4.1: (a) Decay time distribution shape before (blue) and after (green) the selection and (b) the resulting decay time dependent acceptance shape (arbitrary units).

Since the time dependent efficiency $\epsilon(t)$ is not model dependent, it can be derived from simulated data. The correction factor is shown for $\mathsf{B}^0 \to \mu^+\mu^-$ as a function of y_s and $\mathcal{A}_{\Delta\Gamma}$ in Figure 4.2. For the SM ($\mathcal{A}_{\Delta\Gamma}$=1 , y_s=0.0613 $\pm$ 0.0059, $\tau_{\mathsf{B}^0_{s,H}} = 1.618 \pm 0.014\,\text{ps}$) the correction factor is:

$$\delta_\epsilon - 1 \quad = \quad (+4.57 \pm 0.03)\%. \tag{4.3}$$

The uncertainties on this correction are only due to the statistics of the simulated data sample used to compute the decay time dependent efficiency, and they largely cancel since the numerator and denominator in Equation 4.2 are highly correlated.

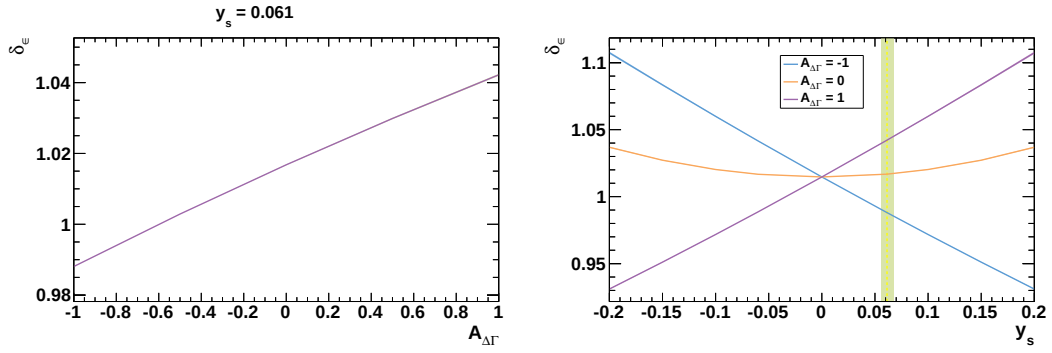
The efficiency for $\mathsf{B}^0 \to \mu^+\mu^-$ was assumed in the published analyses [4, 5, 6, 1] to be identical to that of $\mathsf{B}^0_s \to \mu^+\mu^-$. It should therefore be corrected as well. Since the lifetime value used in the simulation is closer to that for the B^0 $\tau_{\mathsf{B}^0} = (1.519 \pm 0.007)\,\text{ps}$ than for the $\mathsf{B}^0_{s,H}$, the B^0 correction factor is smaller:

$$\delta_\epsilon - 1 \quad = \quad (+1.48 \pm 0.01)\% \tag{4.4}$$

Note that this correction is not model dependent since the width difference in the B^0 mode is negligible.

4.1.3 New Definition of the Normalisation Factor

The efficiency of the inital selection enters the denominator of the normalisation factor defined in Equation 3.3. Therefore the normalisation factor depends on the value of y_s and $\mathcal{A}_{\Delta\Gamma}$. The $\mathsf{B}^0_s \to \mu^+\mu^-$ time integrated branching



(a) (b)

Figure 4.2: (a) Efficencies bias as a function of $\mathcal{A}_{\Delta\Gamma}$ for the experimental y_s value [3] and (b) as a function of y_s for different $\mathcal{A}_{\Delta\Gamma}$ values. The vertical green band is the experimental averaged y_s value and its uncertainty.

fraction prediction depends also on $\mathcal{A}_{\Delta\Gamma}$ and y_s as shown in Equation 1.13 but the initial branching fraction prediction does not. Hence we suggest (following [7, 8, 9]) to redefine the normalisation factor to gather all the $\mathcal{A}_{\Delta\Gamma}$ and y_s dependency:

$$\alpha_{\text{new}} = \mathcal{B}_{\text{norm}} \times \frac{1}{N_{\text{norm}}} \times \frac{f_{\text{norm}}}{f_{\mathsf{B}^0_{(s)}}} \times \frac{\epsilon_{\text{norm}}}{\delta_\epsilon(\mathcal{A}_{\Delta\Gamma}, y_s)\epsilon_{\text{MC}}} \times \frac{1-y_s^2}{1+\mathcal{A}_{\Delta\Gamma}y_s} \quad (4.5)$$

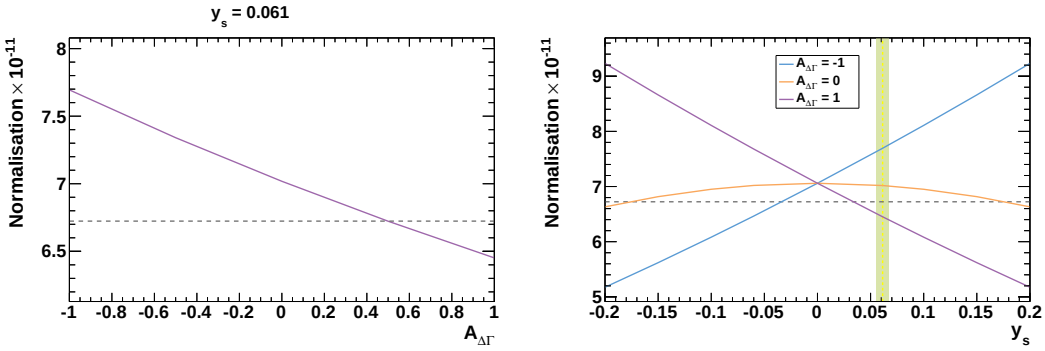
$$= \alpha_{\text{old}} \times \frac{1}{\delta_\epsilon(\mathcal{A}_{\Delta\Gamma}, y_s)} \times \frac{1-y_s^2}{1+\mathcal{A}_{\Delta\Gamma}y_s}. \quad (4.6)$$

Hence the experimental results obtained with this new normalisation factor should be compared to the initial branching fraction.

The new $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ normalisation factor is represented in Figure 4.3 as a function of y_s and $\mathcal{A}_{\Delta\Gamma}$. Note that measured branching fraction is proportional to α (see Equation 3.3), hence the results published in [1] have the same dependency.

4.1.4 Model Dependent BDT PDF

The effect described in Section 4.1.1 also enters the BDT PDF calibration. The BDT uses the candidate decay time as input variable, hence the $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ signal time dependent decay rate in each BDT bin is biased. Were these biases identical in all bins, the PDF would not be affected. However combinatorial background candidates resemble more short-lived than long-lived signal candidate. As combinatorial background candidates are classified



(a) (b)

Figure 4.3: (a) New normalisation factor as a function of $\mathcal{A}_{\Delta\Gamma}$ for the experimental y_s value [3] and (b) as a function of y_s for different $\mathcal{A}_{\Delta\Gamma}$ values. The dashed line represents the normalisation of the 2012 analysis [1] corrected by $(1 - y_s)$ as the correction on the SM prediction due to B^0_s-$\overline{\mathsf{B}}^0_\mathsf{s}$ mixing was accounted for by comparing the results to the SM time integrated branching fraction prediction.

mainly in the low BDT region, the higher the BDT the larger the mean decay time, as shown in Figure 4.4.

In the published analyses, the BDT PDF is obtained with a calibration method which assumes the BDT PDF to be the same for the $\mathsf{B}^0_{(\mathsf{s})} \to \mathsf{h}^+\mathsf{h}'^-$ control channel (cc) and $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$:

$$\frac{N^i_{\rm sig}}{N^{tot}_{\rm sig}} = \frac{N^i_{\rm cc}}{N^{tot}_{\rm cc}}. \tag{4.7}$$

This relation is in fact an approximation of

$$\frac{N^{exp,i}_{\rm sig}}{N^{exp}_{\rm sig}} = \frac{N^{exp,i}_{\rm cc}}{N^{exp}_{\rm cc}} \times \frac{\epsilon^i_{\rm sig}(\mathcal{A}_{\Delta\Gamma}, y_s)}{\epsilon^i_{\rm cc}} \times \frac{\epsilon_{\rm cc}}{\epsilon_{\rm sig}(\mathcal{A}_{\Delta\Gamma}, y_s)}. \tag{4.8}$$

which neglects the two efficiencies ratios. We argue here that these terms must be considered as:

- the signal and the control channels have different time dependent decay rates, hence the efficiency for selecting a BDT bin, ϵ^i, is different for the signal and the control channels: $\epsilon^i_{sig}/\epsilon^i_{cc} \neq 1$,
- the decay time dependent efficiency is different in each BDT bin, hence the ratio $\epsilon^i_{sig}/\epsilon^i_{cc}$ are different and do not cancel in the PDF normalisation,

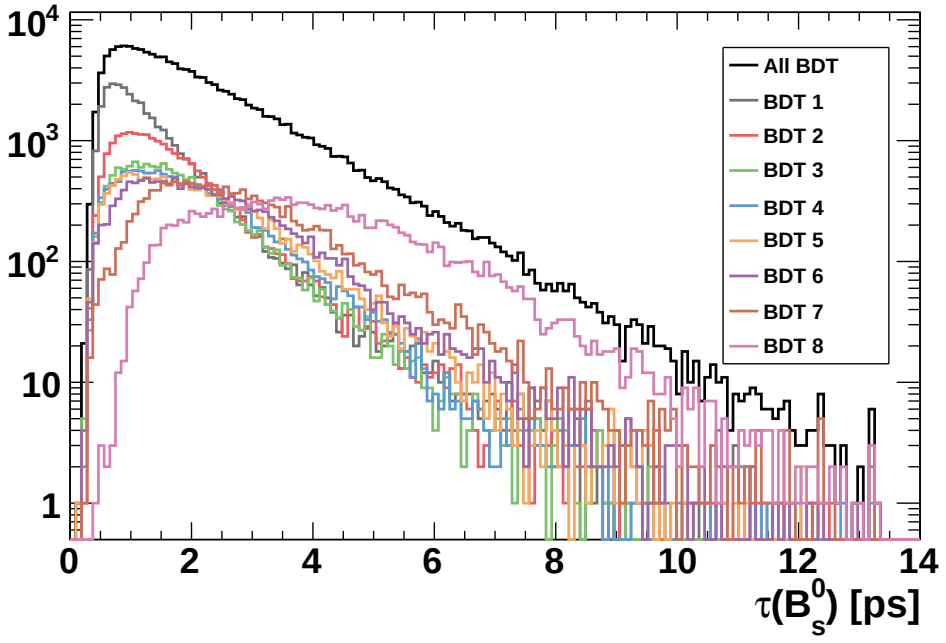


Figure 4.4: Decay time distribution in each BDT bin and in all BDT bins.

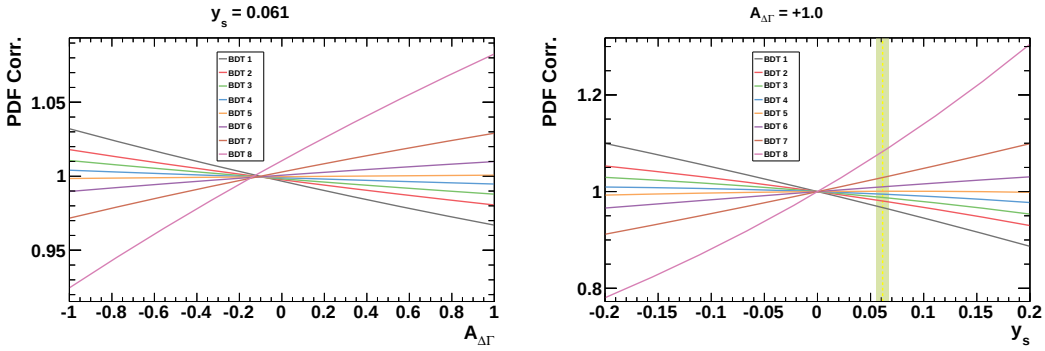
– the signal efficiency is model dependent $\epsilon^i_{sig} = \epsilon^i_{\rm sig}(\mathcal{A}_{\Delta\Gamma}, y_s)$, and so is the PDF.

The correction to be applied to the PDF obtained with Equation 4.7 is

$$\delta^i_{PDF} = \frac{\epsilon^i_{\rm sig}(\mathcal{A}_{\Delta\Gamma}, y_s)}{\epsilon^i_{\rm cc}} \times \frac{\epsilon_{\rm cc}}{\epsilon_{\rm sig}(\mathcal{A}_{\Delta\Gamma}, y_s)}, \tag{4.9}$$

and is shown in Figure 4.5 as a function of $\mathcal{A}_{\Delta\Gamma}$ and y_s for $\mathsf{B}^0 \to \mathsf{h}^+\mathsf{h}'^-$ as control channel. In the analysis, $\mathsf{B}^0_\mathsf{s} \to \mathsf{h}^+\mathsf{h}'^-$ control channels are also used but these modes are three times less abundant than $\mathsf{B}^0 \to \mathsf{h}^+\mathsf{h}'^-$. In addition, the effective lifetime of these modes which are the quantities driving the corrections are $\tau_{\mathsf{B}^0_\mathsf{s}\to\mathsf{K}^+\mathsf{K}^-} = (1.440 \pm 0.009)$ ps [10] and $\tau_{\mathsf{B}^0_\mathsf{s}\to\pi^+\mathsf{K}^-} = \tau_{\mathsf{B}^0_{\mathsf{s,L}}} = (1.431 \pm 0.009)$ ps which is, at maximum, only 5% smaller than $\tau_{\mathsf{B}^0\to\mathsf{h}^+\mathsf{h}'^-} = (1.519 \pm 0.007)$ ps. We therefore neglect them in the PDF correction.

The correction to be applied for the SM is reported for each bin in Table 4.1. The effect of these corrections on the final results cannot be derived directly as each bin has a different sensitivity. Hence Table 4.1 reports also the contribution each bin brings to the total analysis sensitivity [11]. Weight-



(a) (b)

Figure 4.5: (a) Correction to be applied to the BDT PDF as a function of $\mathcal{A}_{\Delta\Gamma}$ for the experimental average y_s value [3] and (b) as a function of y_s for the SM $\mathcal{A}_{\Delta\Gamma}$ value (+1). The vertical green band is the experimental average y_s value and its uncertainty.

ing the corrections by these numbers, the shift of the final branching fraction results due to the PDF can be estimated at 3.9% for the 2012 analysis.

A more correct evaluation of this shift was performed by comparing the fit results with and without the corrections. As shown in Table 4.2, the global shift due to the PDF is estimated at 3.4% and 3.3% for 2012 and the combined 2011-2012 analyses which is consistent with the previous estimate.

For $\mathsf{B}^0 \to \mu^+\mu^-$ no correction has to be applied since the control channel is also a B^0 decay and has therefore the same decay rate as the signal.

4.1.5 Corrected Results

For $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$, the corrections to be applied are model dependent. For the SM, the published analysis results [1] are biased by $(-6.57 \pm 0.03)\%$. Including the additional -6.1% correction required to translate the $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ proper time integrated branching fraction into the initial CP-average one, the total shift is $(-12.30 \pm 0.56)\%$. Therefore the branching fraction measured in [1] which was $\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-) = (3.2^{+1.5}_{-1.2}) \times 10^{-9}$, and compared to $(3.57 \pm 0.30) \times 10^{-9}$ is now for the SM:

$$\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-) \quad = \quad (2.8^{+1.3}_{-1.1}) \times 10^{-9}, \tag{4.10}$$

and is to be compared to $3.34 \pm 0.27 \times 10^{-9}$. For other models, the results are shown as a function of $\mathcal{A}_{\Delta\Gamma}$ in Figure 4.6. For this curve the correction due to the initial selection has been computed for each $\mathcal{A}_{\Delta\Gamma}$ value. However the BDT calibration correction has been evaluated by interpolating linearly

Table 4.1: Corrections to the BDT PDF in each bin in the SM ($\mathcal{A}_{\Delta\Gamma} = 1$, $\tau_{\mathsf{B}^0_\mathsf{s}} = (1.519 \pm 0.007)\,\text{ps}$ and $y_s = 0.0613 \pm 0.0059$) and contribution from each bin to the total sensitivity [11].

Bin	PDF Correction $\delta^i_{PDF} - 1$ (%)	Sensitivity Contribution %
1	-3.30 ± 0.03	0.0981
2	-1.94 ± 0.04	1.09
3	-1.21 ± 0.04	2.03
4	-0.53 ± 0.05	4.33
5	$+0.07 \pm 0.05$	8.70
6	$+0.99 \pm 0.06$	16.1
7	$+5.63 \pm 0.07$	67.6

Table 4.2: Fit results with and without the BDT PDF correction.

	$\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-)$ $[10^{-9}]$		Bias
	Nominal PDF	Corrected PDF	%
2011	$1.35^{+1.67}_{-1.19}$	$1.31^{+1.62}_{-1.15}$	-2.8
2012	$5.05^{+2.44}_{-1.95}$	$4.91^{+2.39}_{-1.91}$	-3.4
2011-12	$3.18^{+1.44}_{-1.19}$	$3.07^{+1.41}_{-1.15}$	-3.3

the corrections at $\mathcal{A}_{\Delta\Gamma} = 1$ (-3.3%) and at $\mathcal{A}_{\Delta\Gamma} = -0.1$ as Figure 4.5 shows that for this point no correction has to be applied. The branching fraction central value varies by $\sim 29\%$ over the $\mathcal{A}_{\Delta\Gamma}$ range. Hence, despite the large experimental error ($\sim 40\%$), the $\mathcal{A}_{\Delta\Gamma}$ dependency is still relevant.

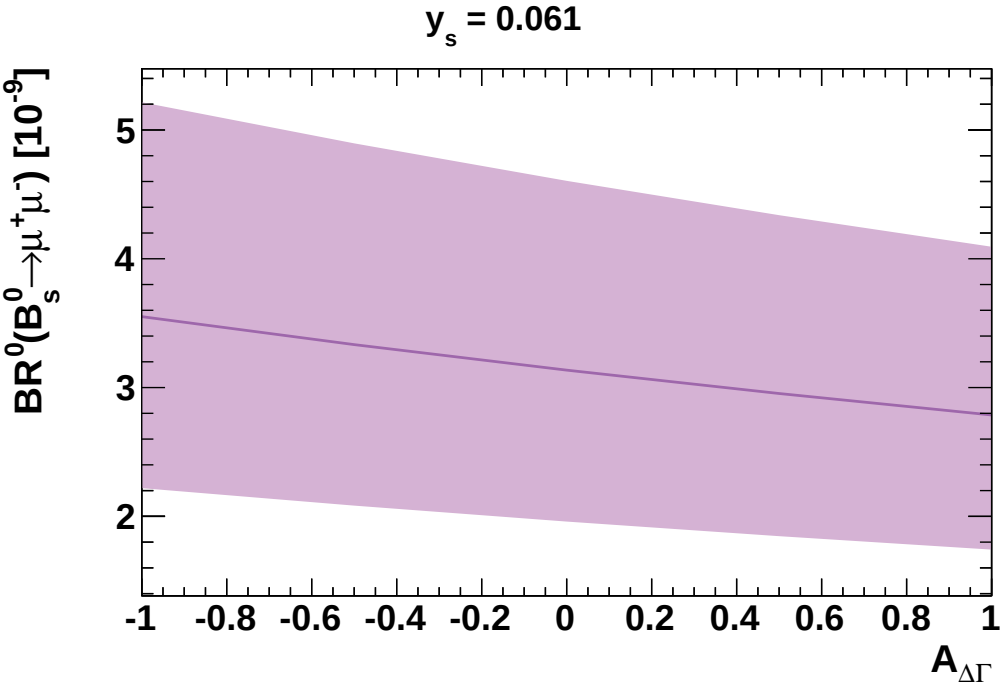


Figure 4.6: $\mathcal{B}(\mathsf{B}^0_s \to \mu^+\mu^-)$ as a function of $\mathcal{A}_{\Delta\Gamma}$ with y_s=0.0613 ± 0.0059 [3].

For $\mathsf{B}^0 \to \mu^+\mu^-$ the correction is smaller and amounts to $(-1.46 \pm 0.01)\%$. Hence, the 95% confidence level upper limits of the latest analysis [1] is shifted from $\mathcal{B}(\mathsf{B}^0 \to \mu^+\mu^-) < 9.4 \times 10^{-9}$ to:

$$\mathcal{B}(\mathsf{B}^0 \to \mu^+\mu^-) \quad < \quad 9.3 \times 10^{-10}. \tag{4.11}$$

4.2 Theoretical Predictions Confronting Experimental Results

Section 4.1 explained that experimental results depend on $\mathcal{A}_{\Delta\Gamma}$. In light of this conclusion, the comparison between experimental results and theoretical predictions found in the literature are to be revisited. This comparison in mainly done in a model independent way in Section 4.2.1. Then, a short discussion on specific models is held in Section 4.2.2.

4.2.1 Model Independent Constraints

The corrected results in Equation 4.10 are in agreement with the SM prediction but constrain physics beyond the SM. These constraints start allowing to disentangle various phenomenologies. Figure 4.7 shows the regions in the plane $\mathcal{B}(\mathrm{B_s^0} \to \mu^+\mu^-) \times \mathcal{A}_{\Delta\Gamma}$ allowed for scalar and non-scalar scenarios superimposed with the previous results[2] and indicates that at 68% C.L. most of the scalar scenario phase space is excluded. Above all, Figure 4.7 demonstrates the importance of comparing prediction and experimental results as functions of $\mathcal{A}_{\Delta\Gamma}$. This statement will be quantified in the following in terms of Wilson coefficient constraints.

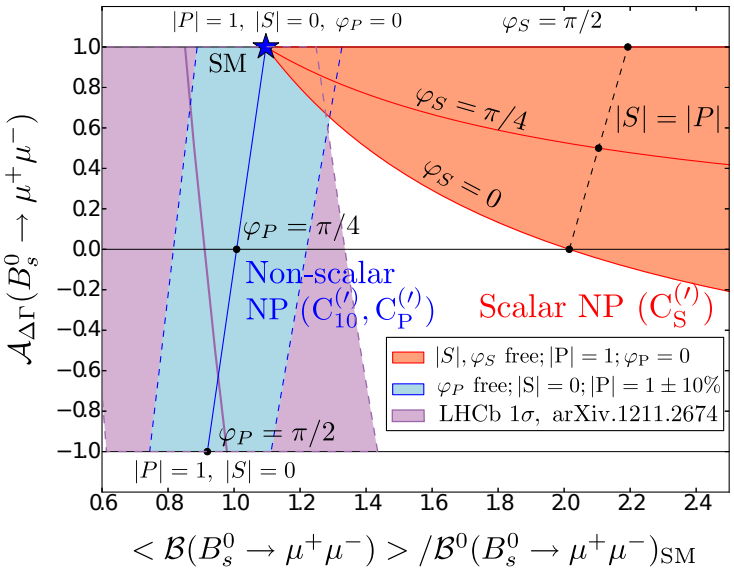
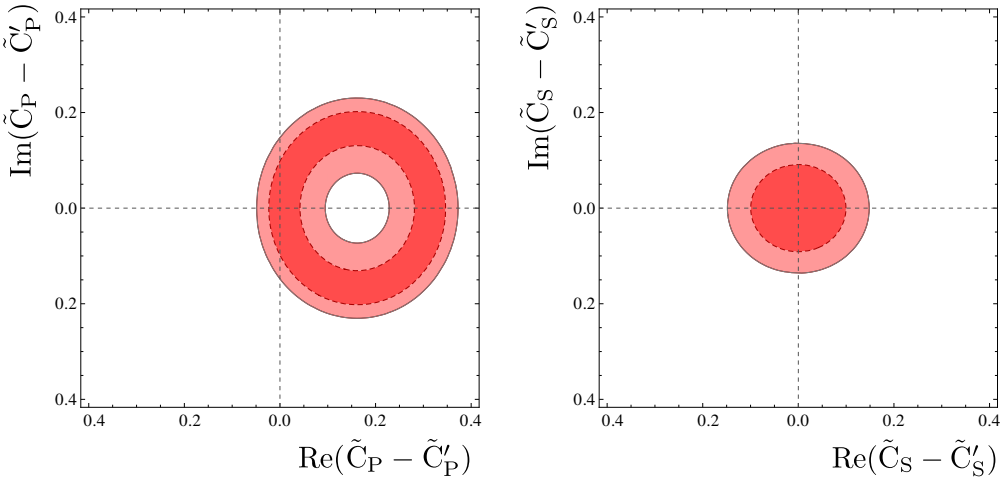


Figure 4.7: Allowed region in the plane $\mathcal{B}(\mathrm{B_s^0} \to \mu^+\mu^-)/\mathcal{B}(\mathrm{B_s^0} \to \mu^+\mu^-)_{\mathrm{SM}} \times \mathcal{A}_{\Delta\Gamma}$ if new physics contributions are scalar or not superimposed with the experimental constraint (modified from [12]). S, P, $\varphi_{S,P}$ are defined in Equation 1.20.

The constraints on the Wilson coefficients shown in Section 1.5 can also be updated [13, 14] with the new results as shown in Figure 4.8 and Figure 4.9. The new external contours are slightly looser than in Figure 1.10 and Figure 1.11, but the internal contours in the $\mathrm{Re}(\mathrm{C_P} - \mathrm{C'_P}) - \mathrm{Im}(\mathrm{C_P} - \mathrm{C'_P})$

[2] In [12], from which is taken the scalar and non-scalar predictions, the values used for y_s and $\tau_{\mathrm{B_s^0}}$ are 0.088 ± 0.014 and 1.466 ± 0.031 ps, hence the results shown in Figure 4.7 have been re-computed with respect to Figure 4.6 with the y_s and $\tau_{\mathrm{B_s^0}}$ values used in [12].

and $\mathrm{Re}(\mathrm{C}_{10}^{(\prime)}) - \mathrm{Im}(\mathrm{C}_{10}^{(\prime)})$ planes are more stringent due to the lower bound the 3.5 σ signal evidence allows to set on the branching fraction.



(a) (b)

Figure 4.8: Values for the $\tilde{C}_P^{(\prime)}$ (a) and $\tilde{C}_S^{(\prime)}$ (b) Wilson coefficients consistent at 1 σ with $\mathcal{B}(\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-) = 3.2^{+1.5}_{-1.2} \times 10^{-9}$ if $C_{10}^{(\prime)}$ is purely SM [13]. $\tilde{C}_{S,P}^{(\prime)}$ are defined in Equation 1.31. The constraints improve slightly with respect to Figure 1.10.

To strengthen our statement on the importance of comparing prediction and experimental results as a function of $\mathcal{A}_{\Delta\Gamma}$ we computed the bound on the plane $\mathrm{Re}(\mathrm{C_S} - \mathrm{C'_S}) - \mathrm{Im}(\mathrm{C_S} - \mathrm{C'_S})$ with and without considering the $\mathcal{A}_{\Delta\Gamma}$ dependency of the results. Considering only new physics in $C_S^{(\prime)}$, $\mathcal{B}(\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-)$ is given by [15]:

$$\frac{< \mathcal{B}(\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-) >}{\mathcal{B}(\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-)_{\mathrm{SM}}} = (1 + |S|^2) \times \frac{1 + \mathcal{A}_{\Delta\Gamma} y_s}{1 - y_s^2}, \tag{4.12}$$

with S defined in Equation 1.20 and $\mathcal{A}_{\Delta\Gamma}$ which simplifies to:

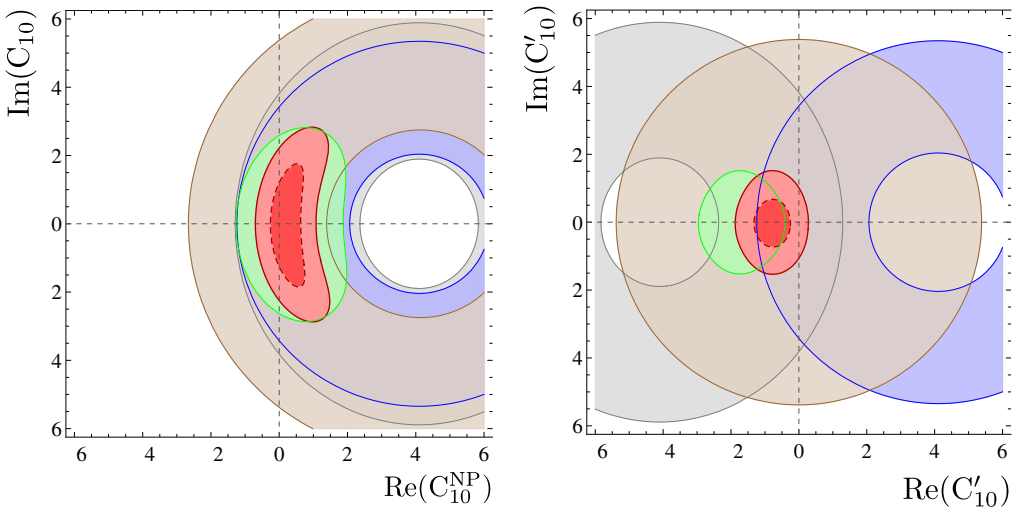
$$\mathcal{A}_{\Delta\Gamma} = \frac{1 - Re(S^2)}{1 + |S|^2}. \tag{4.13}$$

The results we obtained are shown in Figure 4.10. We set the ratio $m_\mathsf{b}/(m_\mathsf{b} + m_\mathsf{s})$ to 1 in S and used the inputs from [7] reported in Table 4.3. The con-

Table 4.3: Inputs used to compute the constraints on C_S, from [7].

$C_{10}^{\rm SM}$	-4.134
m_μ	105.6584 MeV
$m_{\mathsf{B}_\mathsf{s}^0}$	5366.77 MeV

straints without considering the $\mathcal{A}_{\Delta\Gamma}$ dependency are in agreement with [13][3] and become more stringent by considering the $\mathcal{A}_{\Delta\Gamma}$ dependency. Indeed the allowed region in the plane $\mathrm{Re}(\mathrm{C_P} - \mathrm{C'_P}) - \mathrm{Im}(\mathrm{C_P} - \mathrm{C'_P})$ is reduced by approximately 30%.



(a) (b)

Figure 4.9: Values for the C_{10} (a) and C'_{10} (b) Wilson coefficients consistent at 1 σ with the experimental results from $\mathsf{B}^0 \to \mathsf{X_s}\ell^+\ell^-$ (brown), $\mathsf{B}^0 \to \mathsf{K}\mu^+\mu^-$ (blue), $\mathsf{B}^0 \to \mathsf{K}^{*0}\mu^+\mu^-$ (green), $\mathcal{B}(\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-) = 3.2^{+1.5}_{-1.2} \times 10^{-9}$ (gray) and with their combination (red) at 1 and 2 σ [14]. The constraints improve slightly with respect to Figure 1.11.

[3]Some differences are expected as [13] uses $\tau_{\mathsf{B}_\mathsf{s}^0} = (1.466 \pm 0.031)\,\mathrm{ps}$, $y_s = 0.088$ and $C_{10}^{\rm SM} = -4.103$ [16] while we use $\tau_{\mathsf{B}_\mathsf{s}^0} = (1.519 \pm 0.007)\,\mathrm{ps}$, $y_s = 0.0613 \pm 0.0059$, and $C_{10}^{\rm SM} = -4.134$.

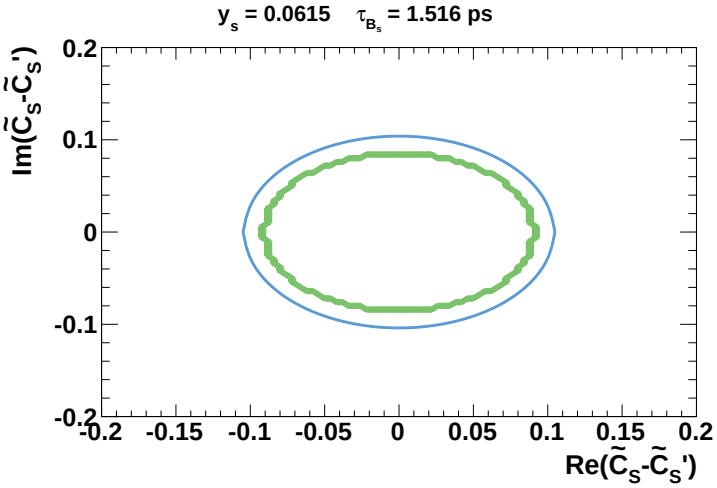


Figure 4.10: Values in the plane $\mathrm{Re}(\tilde{\mathrm{C}}_{\mathrm{S}} - \tilde{\mathrm{C}}'_{\mathrm{S}}) \times \mathrm{Im}(\tilde{\mathrm{C}}_{\mathrm{S}} - \tilde{\mathrm{C}}'_{\mathrm{S}})$ consistent with the experimental results with (green) and without (blue) considering the $\mathcal{A}_{\Delta\Gamma}$ dependency. The allowed regions are the inner part of the ellipses. $\tilde{C}^{(\prime)}_{S,P}$ are defined in Equation 1.31.

4.2.2 Model Dependent Constraints

Model dependent constraints can also be derived with these results. In this case the constraints are applied to the phase space of the parameters of the model considered. Figure 4.11 (modified from [17]) gives an idea on how the model phase space reduces by considering these new results. For an extensive study on the constraints imposed on SUSY parameters by these results the reader is referred to [18]. This study shows for instance that in the CMSSM the Higgs mass measured by the ATLAS and CMS collaborations prevents $\mathcal{B}(\mathrm{B}^0_{\mathrm{s}} \to \mu^+\mu^-)$ to be smaller than the SM prediction.

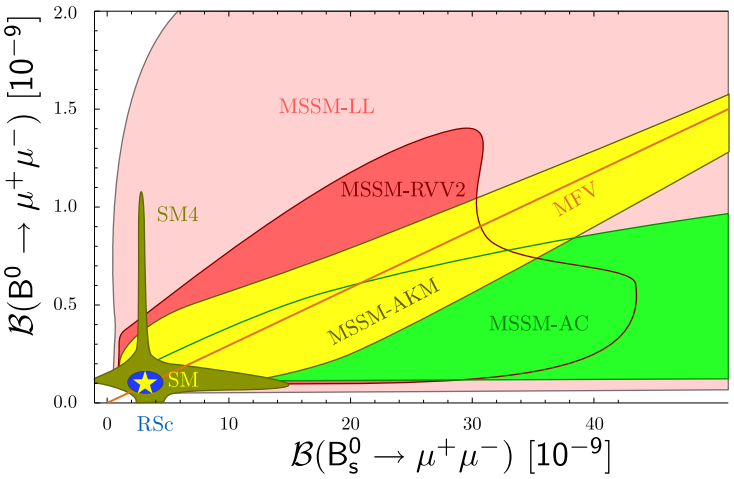


Figure 4.11: Prediction for the branching fractions of $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ in Minimal Flavour Violation models (MFV, see Section 1.5.2), the SM with a fourth generation (SM4, see Section 1.5.2), the Randall-Sundrum model (RSc) with custodial protection [19] and four Minimal SUSY favour models (MSSM), Agashe and Carone (AC) [20], Ross, Velasco-Sevilla and Vives (RVV2) [21], Antusch, King and Malinsky (AKM10) [22] and a model with left-handed currents only (LL11) [23]. The SM prediction is marked by a star [17]. The shaded area is excluded at 95% C.L. by the results from [1].

4.3 Conclusion

This chapter provided a comparison between the experimental results and the theoretical predictions described, respectively, in Section 3 and Section 1. In order to perform this comparison correctly, we studied the model dependency of the experimental results arising from a degree of freedom left in the analysis by the unknown value of $\mathcal{A}_{\Delta\Gamma}$. This dependency is not negligible. Indeed, the measured $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ branching fraction varies by 29% over the $\mathcal{A}_{\Delta\Gamma}$ range while its uncertainty is 40%.

The comparisons found in the literature do not include this effect, but are already reducing sharply the available phase space for new physics. We showed that accounting for the $\mathcal{A}_{\Delta\Gamma}$ dependency of the results allows to improve these constraints. In particular the allowed region in the plane $\mathrm{Re}(\tilde{\mathrm{C}}_\mathrm{S} - \tilde{\mathrm{C}}'_\mathrm{S}) \times \mathrm{Im}(\tilde{\mathrm{C}}_\mathrm{S} - \tilde{\mathrm{C}}'_\mathrm{S})$ is reduced by 30%.

Bibliography

[1] LHCb Collaboration. *First evidence for the decay* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$. Physical Review Letter, **110** (2013), 021801.

[2] M. Perrin-Terrin, et al. *Towards a model independent branching ratio measurement of* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$. LHCb-INT-2013-012 (2013).

[3] O. L. for HFAG. private communication.

[4] LHCb Collaboration. *Search for the rare decays* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *and* $\mathsf{B}^0 \to \mu^+\mu^-$. Physics Letter, **B699** (2011), 330–340.

[5] LHCb Collaboration. *Search for the rare decays* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *and* $\mathsf{B}^0 \to \mu^+\mu^-$. Physics Letter, **B708** (2012), 55–67.

[6] LHCb Collaboration. *Strong constraints on the rare decays* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *and* $\mathsf{B}^0 \to \mu^+\mu^-$. Physical Review Letter, **108** (2012), 231801.

[7] A. Buras, et al. *On the Standard Model prediction for* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$. European Physical Journal, **C72** (2012), 2172.

[8] A. J. Buras, et al. *The Anatomy of Z' and Z with Flavour Changing Neutral Currents in the Flavour Precision Era*. Journal of High Energy Physics, **1302** (2013), 116.

[9] K. D. Bruyn, et al. *Branching Ratio Measurements of* B_s^0 *Decays*. Physical Review, **D86** (2012), 014027.

[10] LHCb Collaboration. *Measurement of the effective* $\mathsf{B}_\mathsf{s}^0 \to \mathsf{K}^+\mathsf{K}^-$ *lifetime*. Physics Letter, **B707** (2012), 349–356.

[11] M. Perrin-Terrin et al. *Optimisation of the binning of the discriminating variables used in the computation of* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *upper limits with the modified frequentist approach*. LHCb-INT-2012-003 (2012).

[12] K. D. Bruyn, et al. *Probing New Physics via the* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *Effective Lifetime*. Physical Review Letter, **109** (2012), 041801.

[13] W. Altmannshofer. *The* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *and* $\mathsf{B}^0 \to \mu^+\mu^-$ *Decays: Standard Model and Beyond*. ArXiv:1306.0022, (2013).

[14] D. M. Straub. *Constraints on new physics from rare (semi-)leptonic B decays*. ArXiv:1305.5704, (2013).

[15] A. J. Buras, et al. *Probing New Physics with the* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *Time-Dependent Rate.* Journal of High Energy Physics, **1307** (2013), 77.

[16] W. Altmannshofer, et al. *Model-Independent Constraints on New Physics in* $\mathsf{b} \to \mathsf{s}$ *Transitions.* Journal of High Energy Physics, **1204** (2012), 008.

[17] D. M. Straub. *Overview of Constraints on New Physics in Rare B Decays.* ArXiv:1205.6094, (2012).

[18] A. Arbey, et al. *Supersymmetry confronts* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *Present and future status.* Physical Review, **D87** (2013), 035026.

[19] M. Blanke, et al. *Rare* K *and* B *Decays in a Warped Extra Dimension with Custodial Protection.* Journal of High Energy Physics, **0903** (2009), 108.

[20] K. Agashe et al. *Supersymmetric flavor models and the* $\mathsf{B} \to \phi\mathsf{K}_\mathsf{S}^0$ *anomaly.* Physical Review, **D68** (2003), 035017.

[21] G. G. Ross, et al. *Spontaneous CP violation and nonAbelian family symmetry in SUSY.* Nuclear Physics, **B692** (2004), 50–82.

[22] S. Antusch, et al. *Solving the SUSY Flavour and CP Problems with SU(3) Family Symmetry.* Journal of High Energy Physics, **0806** (2008), 068.

[23] L. J. Hall et al. *A Geometry of the generations.* Physical Review Letter, **75** (1995), 3985–3988.

Chapter 5

An Inclusive Event Vertex Reconstruction

The $B^0_{(s)} \to \tau^+\tau^-$ channels are also interesting probes of new physics and are complementary to the $B^0_{(s)} \to \mu^+\mu^-$ modes. However these channels are challenging experimentally as the τ needs to be reconstructed from its decay products which necessarily contain at least one neutrino. This chapter introduces a reconstruction method based on a function returning at each point the probability to have a vertex. General motivations to develop this method are discussed in Section 5.1. Then, Section 5.2 describes the reconstruction algorithm. Section 5.3 shows that, in order to operate correctly in the LHCb environment, some sub-algorithms must be refined. The algorithm application to $B^0_{(s)} \to \tau^+\tau^-$ is presented in Section 5.4 with a comparison to the classical reconstruction based on combinatorial trials. Finally, Section 5.5 shows that the realm of application of this method is wider than $B^0_s \to \tau^+\tau^-$ and introduces some new isolation variables to separate the $B^0_s \to \mu^+\mu^-$ signal from its combinatorial background.

5.1 Motivations

In most LHCb analyses, signal candidates are reconstructed with a combinatorial method. First, sets of tracks are preselected based on the features of the decay products. Then all combinations of tracks from these sets are tried and the best candidates are retained.

In experiments installed on e^+e^- machines, it was possible to conceive more refined candidate reconstruction, as the events contain less tracks. In particular in SLD (SLAC Large Detector) a technique was developed [1] to reconstruct candidates with a global function taking a list of tracks as input

and returning a probability at each point to have a vertex. For a review of the vertexing techniques, the reader is referred to [2].

In LHCb, due to the very large number of tracks in the event, implementing such a method is very challenging. However, the challenge is worthwhile, especially for sophisticated modes where the tracks in the final states do not have very specific kinematic features per se, but their configuration (isolation for example) with respect to the other tracks in the event might. In addition inclusive techniques might be beneficial to flavour tagging algorithms.

The SLD algorithm, called ZVTOP [1], was originally coded in Fortran and was recently re-encoded in the C++ package LCFIVertex [3] developed for the ILC (International Linear Collider) project.

Our first attempt at implementing the method in the LHCb software was to use this package providing adapters to convert the LHCb linear track parametrisation into the LCFI helicoidal coordinate system. After a proof of principle, it was judged more reliable and efficient to encode the method in the LHCb software from scratch. A first Python coded prototype was developed and, once the package architecture was established, the algorithm was written in C++.

5.2 General Algorithm

The inclusive event vertex reconstruction algorithm was implemented in the LHCb software on the basis of the LCFIVertex implementation [3]. The key element of the method is the vertex function which is described in Section 5.2.1. The algorithm in itself is then explained in Section 5.2.2. Finally the algorithm parameters are listed in Section 5.2.3.

5.2.1 Vertex Function

Gaussian Tube

A reconstructed track is a trajectory with an error or equivalently a tube in which we know at a given C.L. that a real track passed. The probability for the real track to have passed by a given point $\mathbf{r}$ is:

$$f_i(\mathbf{r}) \;=\; \exp\{-\frac{1}{2}(\mathbf{r}-\mathbf{p})\mathbb{V}^{-1}(\mathbf{r}-\mathbf{p})^T\}, \tag{5.1}$$

where $\mathbf{p}$ is the point of closest approach on the track to $\mathbf{r}$ and $\mathbb{V}$ is the 3D covariance matrix of the track at $\mathbf{p}$. Thus, in the plane transverse to the track momentum, the probability is distributed as a Gaussian function. The tubes are therefore called Gaussian tubes.

Vertex Function

Intuitively, if all Gaussian tube probabilities in an event were added together, the resulting function should feature peaks where tracks encounter i.e. at the vertex positions. Finding vertices is then equivalent to find the maxima of the function.

In practice the vertex function is defined by:

$$V(\mathbf{r}) \;=\; \sum_{i=1}^{N} f_i(\mathbf{r}) - \frac{\sum_{i=1}^{N} f_i^2(\mathbf{r})}{\sum_{i=1}^{N} f_i(\mathbf{r})}, \tag{5.2}$$

so that the function has non null values only where more than two tracks overlap and the more tracks encounter at a point the larger the vertex function. For example in a region where three tracks overlap the vertex function is close to 2.

Resolving Vertices

The maxima of the vertex function define vertex candidates. When more than three tracks originate from a real vertex it can happen that several local maxima are found around the real vertex position. Intuitively we understand that these vertices belong in fact to the same meta-peak. To use this idea in the code, a criterium is defined, so that two vertices are resolved only if:

$$\frac{\min\{V(\mathbf{r}) : \mathbf{r} \in \mathbf{r_1} + \alpha(\mathbf{r_2} - \mathbf{r_1}), 0 \leq \alpha \leq 1\}}{\min\{V(\mathbf{r_1}), V(\mathbf{r_2})\}} \;<\; R_0, \tag{5.3}$$

where R_0 is a threshold parameter. In other words, two peaks correspond to two different vertices if the valley between them is deep enough, as shown in Figure 5.1.

5.2.2 Algorithm

The algorithm consists of six sequential steps described below. Note that after Step 3 a vertex has two vertex function values: V_f, evaluated at the vertex fitted position and V_m, the vertex function maximum closest to the fitted position.

Step 1 All possible pairs of tracks are fitted into vertices. These two-prong vertices are kept only if the χ^2 of the vertex fit is smaller than a threshold χ_0^2 and the vertex function at the fitted position is larger than a threshold V_0. Note that after this step, a track can belong to many vertices.

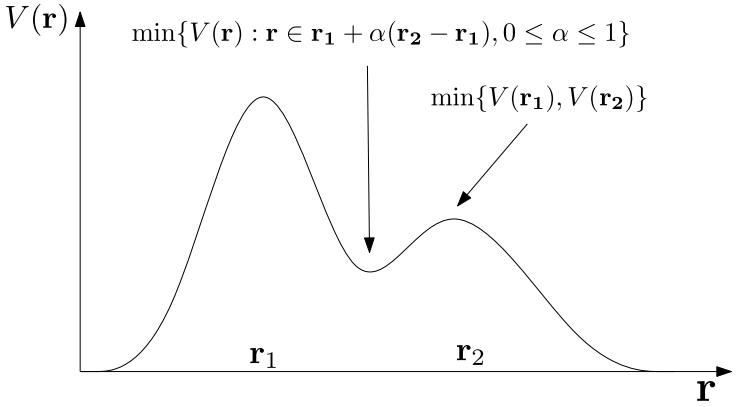


Figure 5.1: Resolution criteria (see Equation 5.3).

Step 2 During this step the algorithm loops over the tracks and performs mainly a cleaning operation. For each track, the list of vertices to which it belongs is built and ordered by decreasing V_f order. The vertex with the largest vertex function V_f^L is therefore in the first position. The track is then removed from all vertices with a vertex function smaller than $\mathcal{F} \times V_f^L$, where $\mathcal{F}$ is a parameter of the algorithm.
The vertices still containing the track are considered by decreasing V_f order and the track is removed from all vertices unresolved from any previously considered vertex. Hence, if two unresolved vertices contain the same track, only the one with the largest vertex function keeps it.

Step 3 All empty vertices are removed. For each vertex the maximum of the vertex function closest to the vertex fitted position is searched and the list of remaining vertices is ordered by decreasing V_m.

Step 4 In this step unresolved vertices are merged into clusters. Vertices are considered by decreasing V_m order and the clustering process is performed iteratively. First all vertices unresolved from the vertex with the highest V_m are merged into the cluster. Then the cluster absorbs all vertices unresolved from any vertex it already contains. The process is iterated until no more vertices are added. Then all clustered vertices are removed for the initial vertex list and the process is repeated.

Step 5 During this step clusters are fitted into vertices.

Step 6 During this step the final vertices are formed. All tracks in the vertex having an impact parameter significance with respect to the vertex larger than a threshold χ^2_{TRIM} are flagged. Then a new vertex is

fitted without these tracks and the process is repeated until no track is flagged. Finally when tracks belong to several vertices, they are attributed to the vertex with the largest V_m.

5.2.3 Algorithm Parameters

In the algorithm described above five parameters have been defined. Their settings are not independent as initial steps have great impacts on later ones. Their main properties are described in the following.

χ_0^2 **and** V_0 define the numbers of two prong vertices passing Step 1. They are critical parameters for the CPU consumption. Higher χ_0^2 and lower V_0 values increase the number of vertices passing Step 1.

R_0 defines when vertices are unresolved. The larger R_0 the more close vertices are merged (Step 5) and the less a track can belong to several vertices (Step 2).

$\mathcal{F}$ controls, together with R_0, the number of vertices a tracks can belong to in the first algorithm steps. The lower $\mathcal{F}$ the more a track can belong to several vertices.

$\chi^2_{\mathbf{TRIM}}$ controls the final vertex quality and so the number of tracks that are rejected to form these vertices. The larger χ^2_{TRIM}, the less tracks are rejected.

5.3 Maximum Finder Algorithms

The reconstruction algorithm presented in the previous section uses a sub-algorithm to find the maxima of the vertex function. Section 5.3.1 shows that the maximum finder used in the original code is not suited for the LHCb environment. Thus, new maximum finders were developed. They use a common tool to evaluate the distance to walk. This tool is presented in Section 5.3.2. Then the core of the maximum finders, i.e. the evaluation of the direction in which to walk, are presented in Section 5.3.3 and the algorithms' performances are compared.

5.3.1 LHCb Environment

One of the key elements for the algorithm is the sub-algorithm used to find the maximum of the vertex function. Indeed running the algorithm with a

profiling tool [4] shows that more than 90% of the CPU time is spent to find maxima. In the algorithm implementation we started from [3] the maxima were sought by stepping along each axis until the maximum was found. This method is extremely slow, CPU consuming, and impracticable in the busy LHCb hadronic environment. In addition, due to the boost of the B mesons in LHCb, the vertex function is very elongated in the direction of the B momentum. As a results, the steps in the transverse and longitudinal directions with respect to the B momentum cannot be equally sized. Indeed if the transverse steps are too large, the algorithm misses the vertex function variation and if the longitudinal steps are too small the algorithm cannot converge in a reasonable time. This situation is well illustrated in Figure 5.2, which shows a projection of the vertex function for a $B_s^0 \rightarrow \tau^+\tau^-$ event with both τ decaying to three charged pions and a neutrino. In this Figure, one τ is associated to a vertex function bump whose projection is an ellipse with minor and major axes of around $50\mu m$ and $600\mu m$. As a conclusion, a maximum finder algorithm adapted to the LHCb environment is required. The maximum finder algorithm can be improved by two means. First an adaptive stepping allowing to make steps wider at the beginning and finer around the maximum would improve the algorithm speed and precision. Then a more precise evaluation of the direction in which to step would improve the algorithm speed. Based on these considerations, three algorithms were developed called, Memory Maximum Finder, Star Maximum Finder, and Minuit Based Maximum Finder. The first two ones use a common tool to estimate the position of the maximum once they have found the direction in which to step.

5.3.2 Parabolic Extrapolation

An adaptive stepping method was developed based on the original SLD software. The idea is to compute the vertex function at two points, one before and one after the current position in the direction of the largest positive slope. These two points together with the current position define a parabola. If we consider this parabola as an approximation of the profile of the vertex function in the direction of the largest positive slope, then the position of the real maximum should be close to the maximum of the parabola. The parabolic extrapolator therefore returns a step which brings the current position to the position at the maximum of the parabola.

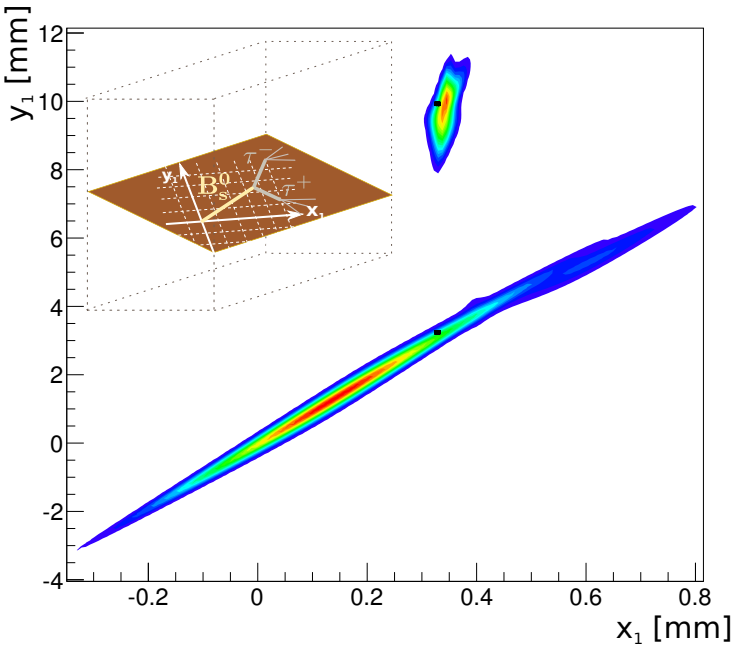


Figure 5.2: Vertex function obtained with the six pions resulting from a simulated $B_s^0 \to \tau^+\tau^-$ decay. The vertex function is shown projected in the decay plane. The origin is set at the B_s^0 true origin vertex and the y axis direction is defined by the true position of the two τ decay vertices. The two black points represent the true projected τ vertex positions.

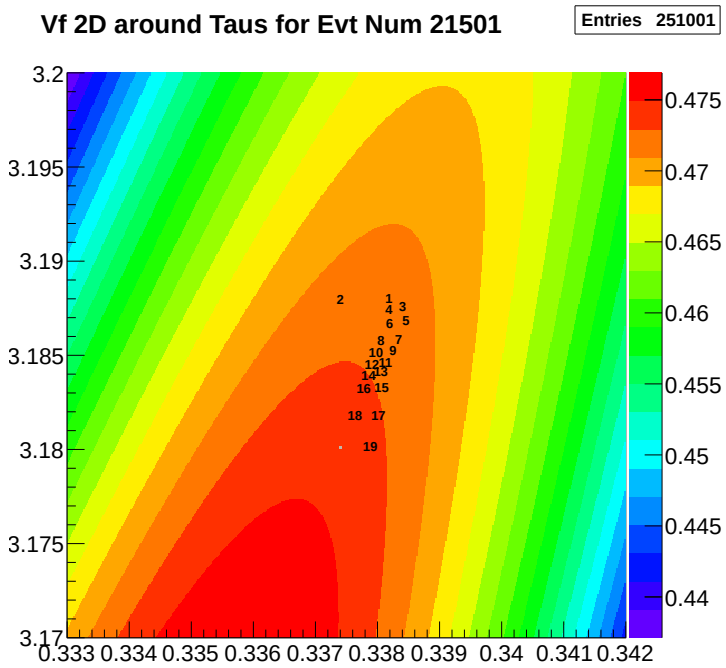


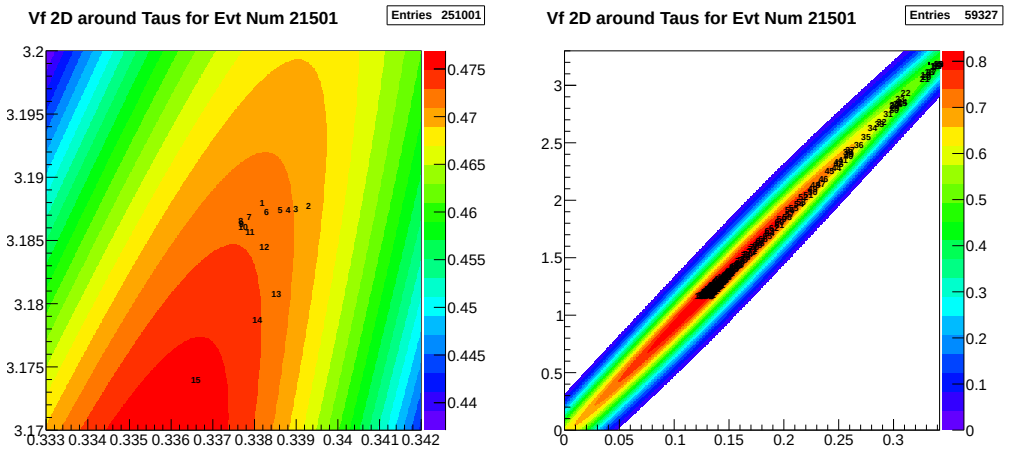
Figure 5.3: Zoom around the initial position of the lower vertex in Figure 5.2. The walk of the algorithm is shown with the black indices.

5.3.3 Step Direction Evaluation

Memory Maximum Finder

Due to the elongated shape of the bumps of the vertex function illustrated in Figure 5.2, the vertex function variation along the direction perpendicular to the crest are very important so the algorithm steps mostly in this direction from one side of the crest to the other and do not progress towards the maximum as shown in Figure 5.3 for the same event as in Figure 5.2.

Assuming that the vertex function shape is approximately symmetric in the direction transverse to the crest, the sum of two consecutive gradient should point in the crest direction and help the algorithm to converge faster. Indeed, as shown in Figure 5.4, when memorising the direction at the previous step, the algorithm converges faster. Note that for this algorithm the gradient of the vertex function was computed analytically by differentiating Equation 5.2.



(a) (b)

Figure 5.4: Zoom around the initial position of the lower vertex in Figure 5.2. The 20 (a) and 200 (b) first steps of the algorithm are shown by the black indices.

Star Maximum Finder

Another algorithm was developed which tries twenty six different directions disposed as to form a regular polyhedra. The algorithm walks in the direction in which the vertex function increases the most by a distance obtained with the parabolic extrapolator.

Minuit Based Maximum Finder

Finally, in the literature [5] the problem of finding a maximum on an ridge is very well understood and solved with conjugate gradient methods. Such methods are already implemented in Minuit [6], we therefore adapted the code to work with this software.

Maximum Finder Algorithms Comparison

Since the maximum of the vertex function may not agree with the position of the true τ decay vertex, it is impossible to know the correct maximum the algorithm should have found. Hence the only solution to check the algorithms' convergence is to compare their results. This comparison was performed with simulated $B_s^0 \to \tau^+\tau^-$ data where the τ decays into three charged pions and a neutrino. The algorithms are initialised at the positions of vertex fitted

with a pair of pion tracks coming from the same τ, hence there is for sure a maximum close to this position. The three algorithms are asked to find this closest maximum. Then the maximum vertex function and maximum position they found are compared as well as the CPU time they use. Figure 5.5 shows that the three algorithms return similar results. However, for some vertices the maxima found by the memory algorithm are higher than those found by Minuit which are already higher than the one found by the star algorithm. To further check the compatibility between the results, the distance between the position at the found maxima are compared in Figure 5.6. This comparison shows that the Minuit and the star algorithm are the most compatible. In average they return the same position within $70\mu m$ while the memory algorithm agrees with the position return by the two other algorithm at more than $160\mu m$. These comparison casts doubts on the convergence of the memory algorithm. The performances of the three algorithms must also be compared in terms of computing time consumption. Figure 5.7 demonstrates that the star and the Minuit algorithm converge faster than the memory one, and the star is even faster than the Minuit algorithm. Hence we decided to use the star algorithm.

5.4 $\mathsf{B}^0_{(\mathsf{s})} \to \tau^+\tau^-$ Reconstruction

The implementation in the LHCb software of the inclusive event vertex reconstruction was primarily motivated by the search for $\mathsf{B}^0_\mathsf{s} \to \tau^+\tau^-$. This channel is described in Section 5.4.1 together with the acceptance, trigger, and signal decay products' reconstruction efficiencies. Then, in Section 5.4.2, a preliminary study on the feasibility of a reconstruction based on the inclusive method is presented and its performances are compared to those obtained with a classical reconstruction based on combinatorial trials. This study is preliminary in the sense that, only the signal efficiency is considered regardless of the global event rate or CPU consumption which are, in fact, major constraints for the reconstruction to be run on large data samples. However, studying and optimising these aspects is worthwhile only after having established the possible improvement the inclusive event vertex reconstruction could bring.

5.4.1 Signal Description

Tau particles being short lived, $\tau_\tau = (0.2906 \pm 0.0010)\,\text{ps}$, they have to be reconstructed from their decay products which, unavoidably, contain at least one neutrino. In order to reconstruct the position of the two τ decay vertices,

we consider the decay modes with three charged tracks[1]. The most abundant three charged body τ decay is $\tau^- \to \pi^-\pi^-\pi^+\nu_\tau$ which has a branching fraction of $(9.31 \pm 0.06)\%$. With this τ decay mode, the effective branching fraction, $\mathsf{B}_\mathsf{s}^0 \to \tau^+(\pi^+\pi^+\pi^-\bar{\nu}_\tau)\tau^-(\pi^-\pi^-\pi^+\nu_\tau)$, is $(6.55 \pm 0.54)\% \times 10^{-9}$, which, as a rule of thumb, is twice as large as the $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ branching fraction.

The acceptance, reconstruction, and trigger efficiencies for these modes can be evaluated from simulations. The generator efficiency is obtained with a sample of simulated data generated with all $\mathsf{B}_\mathsf{s}^0 \to \tau^+\tau^-$ decays products, i.e. the six pions from the two τ decays, in the LHCb acceptance. The efficiency of this generator selection, 14.03%, provides a good estimate of the acceptance efficiency. More up to date and larger simulated data samples are also available. However, these data are produced with some of the candidate selection, also called stripping criteria (pions tracks and B candidates must have momenta and transverse momenta larger than 250 MeV/c and 2 GeV/c), applied directly at the generator level in order to save CPU resources. This generator level selection, together with the requirement of having the six pions in the detector acceptance, have a global efficiency of 2.85%. Hence this simulated data sample is only used to evaluate the reconstruction and trigger efficiencies. Among the produced events passing the generator level selection, 14.33% have the six pions from the two τ reconstructed as long track and the trigger (L0 and HLT) selects 16.00% of them. These efficiencies and the numbers used to compute them are reported in Table 5.2.

5.4.2 Signal Candidate Reconstruction

The standard selection, called stripping, is performed in two steps. First, τ candidates are built from combinations of preselected tracks. Then B candidates are made from combinations of selected τ and filtered out to keep the best candidates. The track's, τ, and B candidates' selection criteria are optimised in order to maximise the signal event rate while respecting CPU time and global event rate limitations imposed by the not infinite computing resources available to run the stripping. This optimisation allows to reach a global efficiency of 0.01440% as shown in Table 5.2. Note that the two simulated data samples, in addition of being generated with different requirements, are stripped differently. Indeed the stripping was re-optimised for the second simulated data sample. The efficiency is given for this re-optimised stripping.

Without considering the CPU time and global event rate constraints mentioned above, we performed a study, on simulated data, to understand

[1]More tracks in the final state would imply strong loss in track reconstruction efficiency.

Table 5.1: Optimal parameter values for the $\mathsf{B}_\mathsf{s}^0 \to \tau^+\tau^-$ reconstruction run on long tracks with an impact parameter with respect to any primary vertex larger than 11 in units of significance.

Parameter	Optimal Value
χ^2_0	1.5
V_0	0.4
$\mathcal{F}$	0.7
R_0	0.4
χ^2_{TRIM}	12

whether or not the inclusive event vertex reconstruction allows, at all, to improve the candidate selection. However, in order to keep the CPU time down to a working level, the tracks on which the reconstruction algorithm is run are passed through a preselection removing tracks coming from the primary vertices. Although the initial motivation was to limit the CPU time, this preselection appeared to be beneficial for the signal reconstruction performances. The figure of merit used to optimise these selection criteria is the number of events for which two vertices each made of only the three pions from the same τ are found. The optimal preselection obtained requires tracks to be of long type and to have impact parameters with respect to any primary vertex in the event larger than 11 in units of significance squared.

The same figure of merit was used to optimise the algorithm parameters whose optimal values are reported in Table 5.1. The parameters were optimised one after another in the order they appear in Table 5.1. Since these parameters are not independent, a few more iterations or, more refined optimisation methods accounting for these dependency, could allow to obtain better performances.

With these set of parameters, the efficiency to reconstruct the signal candidate as two separated vertex (one for each τ) or as a global vertex is 0.01747%. These efficiencies are obtained on simulated data without the tight generation selection. Indeed, this selection being part of the stripping, it must be excluded from this efficiency to correctly estimate the best performances the inclusive reconstruction could reach on presently stripped data.

In summary, the inclusive event vertex reconstruction could improve the global signal efficiency by 21.3% with respect to a classical reconstruction based on combinatorial trial. The global signal efficiencies would then be 0.01747%. This result call for further studies to compare these two ap-

proaches. The inclusive approach performances could be further improved by a better parameter choice or even deeper modifications in the algorithm to better account for the LHCb environment. These optimisations should also account for the CPU and global event rate stripping limitations. Criteria to select candidates out of the vertex list returned by the inclusive algorithm have also to be devised and the treatment of the 6 tracks vertices must be clarified. In any case, the vertex function will provide also useful information to separate between signal and background. The next section demonstrates this point in the case of $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$.

5.5 Isolation Tool

The vertex function used in the reconstruction algorithm contains information on the whole event. Hence it can be a powerful tool to separate signal from background. This section studies the case of the separation of the $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ signal from its combinatorial background. This background is described in Section 5.5.1. One of the main features of the signal with respect to this background is the isolation of the muons from other tracks. Based on this consideration, isolation variables built with the vertex function are introduced and compared to the existing ones in Section 5.5.2.

5.5.1 $\mathsf{B}^0_{(\mathsf{s})} \to \mu^+\mu^-$ Combinatorial Background

Background events in the $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ analysis are mainly of combinatorial origin. The B candidate is formed with muons coming from two different B decays. This type of events and the genuine signal have different topology as shown in Figure 5.8. In particular, muon tracks and B candidate are more isolated from other tracks for signal than for background . These isolation features are seized with two variables Iso_μ and Iso_B defined in Appendix A and shown for signal and background in Figure 5.9. These variables have an excellent discrimination power.

The vertex function provides alternative ways to measure the isolation of the candidate and its muon tracks.

5.5.2 Vertex Function Based Isolation Variables

The typical vertex function profiles obtained by scanning along the muon tracks are shown for signal and background in the top plots of Figure 5.10. Only `Long` type tracks with impact parameter significance with respect to any primary vertex larger than 3 are used in the vertex function. At the

Table 5.2: Number of reconstructed, triggered, and selected candidates and the corresponding efficiencies obtained with a simulation requiring only to have the six pions from the τ in the detector acceptance and with another requiring additional selection criteria. The two last row correspond to the good candidates returned by the topological reconstruction. 3-3 type candidates are candidates where τ are reconstructed as separated vertices and 6 type candidates where the B mesons are reconstructed as a single vertex. The efficiency of the preselection for the inclusive method is included in each of the 3-3 and 6 candidates reconstruction efficiency.

Generation Selection	Pions in Acceptance		Tight Selection	
Stripping Version	17		20	
	Nb Events	Efficiency (%)	Nb Events	Efficiency (%)
Generated	109,999	14.03	3,007,982	2.85
& Reconstructed	7,495	6.814	431,028	14.33
& Triggered L0	2,541	33.90	172,727	40.07
& Triggered HLT1	1,432	56.36	115,853	67.07
& Triggered HLT2	716	50.00	68,939	59.51
& Stripped	143	19.97	18,967	27.51
& Signal Matched	88	61.53	15,202	80.15
Total	–	0.01122	–	**0.01440**
& 3-3 Candidate	50	6.98	5,748	8.338
& 6 Candidate	87	12.2	8,131	11.79
Total	–	**0.01747**	–	0.01315

candidate position, a clear and narrow peak appears which is removed as soon as one of the two muon tracks is removed. For signal no other peaks are expected, while for background peaks are expected downstream to the candidate position where the actual B decays happened. From these different profiles, some V_f based isolation variables can be built.

B Isolation

The B isolation can be quantified by the vertex function at the candidate vertex position $V_{f,\mathrm{B}}$. For the signal $V_{f,\mathrm{B}}$ is expected to be 1 or slightly less, depending on how well the two muon tracks cross. For the background, $V_{f,\mathrm{B}}$ is expected to be more spread around 1, in particular towards higher values due to the other tracks from the two B decays. The signal and background $V_{f,\mathrm{B}}$ distributions are shown in Figure 5.11.

Muon Isolation

The muon isolation is quantified by scanning along the track a vertex function $V_{f,\mu\mu}$ built without the other muon track. As B vertices are expected downstream to the vertex candidate, the scan spans from 3 cm before to 6 cm after the B vertex position. For the signal, in the absence of the other muon, the V_f should be null while for the background peaks are expected downstream to the candidate position. The two signal track isolation variables are combined by taking the highest of the two V_f maxima. This variable is shown in Figure 5.12 together with the position of this maximum. The number of peaks along the scan is also discriminant as it is larger for the background than for the signal and is shown in Figure 5.13.

The standard isolation variable shown in Figure 5.9 is ranked by TMVA[2] as the most discriminating variable. Comparing this variable shown in Figure 5.9 with the V_f based isolation variables shown in Figure 5.11, 5.12, and 5.13 indicates that these new variables can bring a significant discrimination power. In addition the scatter plots in Figure 5.14 demonstrate that the standard and the V_f based isolation variables are mainly uncorrelated, the most correlated ones being the number of $V_{f,\mu\mu}$ peaks and the standard muon isolation. Thus the V_f based isolation variables bring new information and significant discrimination power improvement is foreseen.

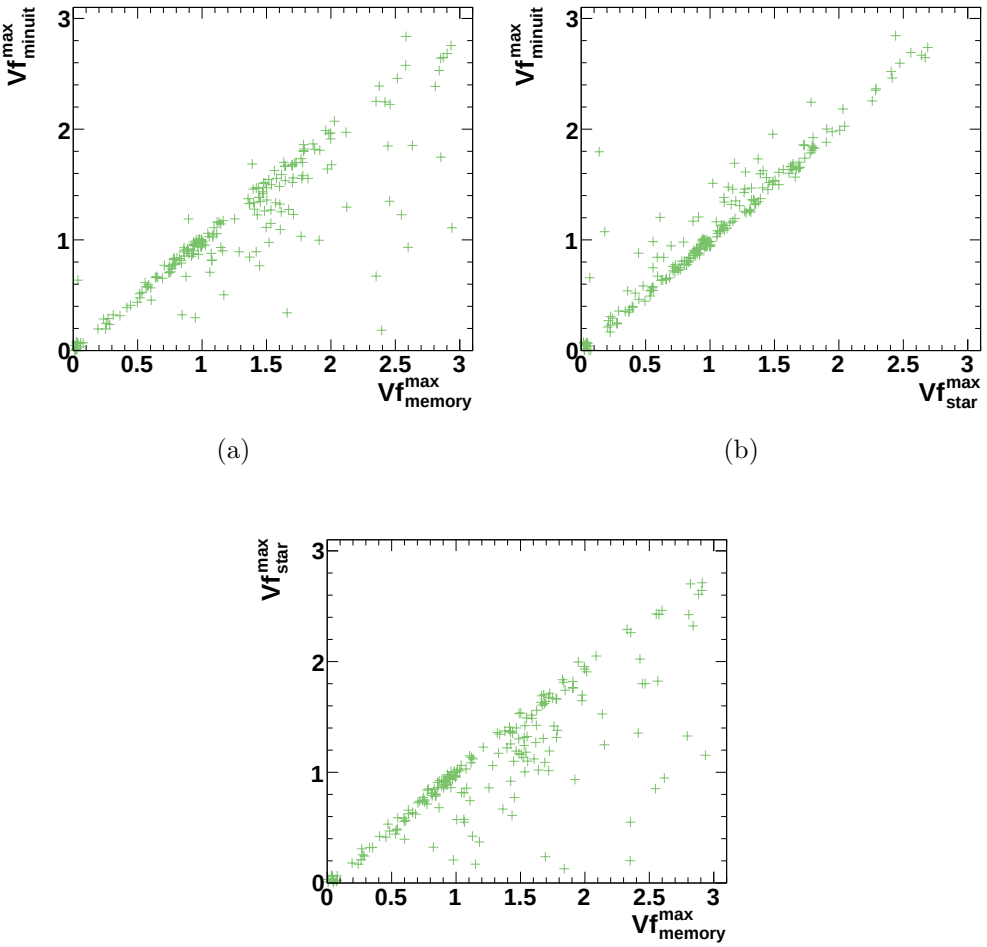
[2]TMVA is a ROOT based software used to build the Boosted Decision Tree in the $\mathrm{B}^0_{(\mathrm{s})} \to \mu^+\mu^-$ analysis (see Section 3.1.5).

5.6 Conclusion

This chapter introduced an original approach to reconstruct vertices based on a global function giving, at each point, the probability to have a vertex. After introducing the motivations for such a reconstruction method, the algorithm has been described in details. Its implementation for the LHCb environment was developed on the basis of the International Linear Collider code. However, some sub-algorithms had to be adapted to the specificities of the LHCb physical environment.

With this adapted algorithm, a preliminary study of the feasibility of a $\mathsf{B}_\mathsf{s}^0 \to \tau^+\tau^-$ analysis based on this method was presented. Compared to a classical reconstruction based on combinatorial trials, the inclusive method could improve the overall signal efficiency up to 21%. This result is promising and calls for other studies to understand whether or not a $\mathsf{B}_\mathsf{s}^0 \to \tau^+\tau^-$ analysis based on this method is feasible.

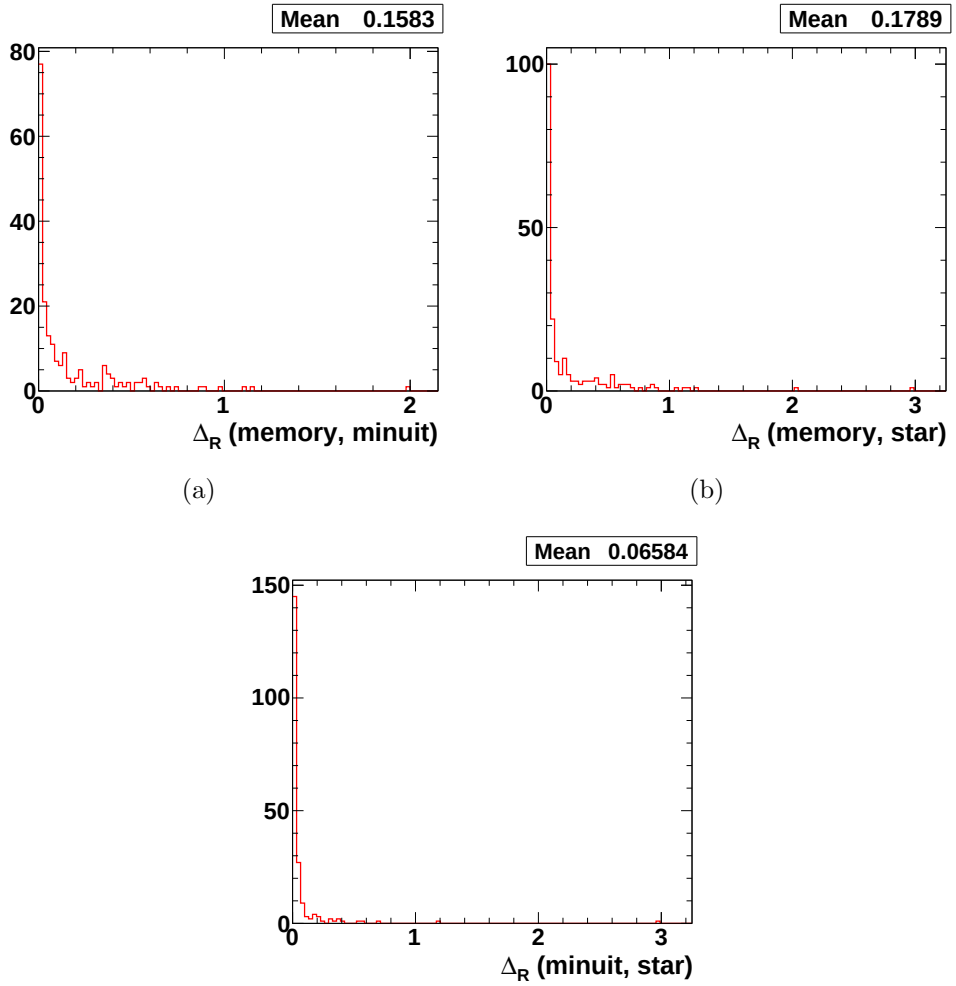
Finally, the range of applications of the topological inclusive event vertex reconstruction is much wider than the search for $\mathsf{B}_\mathsf{s}^0 \to \tau^+\tau^-$. For instance, the vertex function can be used to provide information to separate signals from backgrounds. In particular, we showed that isolation variables can be devised based on the vertex function to separate the $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ signal from its combinatorial background. These new variables show a discrimination power comparable to the best analysis' variables without being correlated to them. Further studies are needed to estimate the gain in sensitivity these variables bring.



(a) (b)

(c)

Figure 5.5: Comparison of the vertex function maximum value found by the Minuit and Memory (a), Star and Memory (b), and Star and Minuit (c) algorithms.



(a) (b)

(c)

Figure 5.6: Distance between the position of the maxima found by the Minuit and Memory (a), Star and Memory (b), and Star and Minuit (c) algorithms.

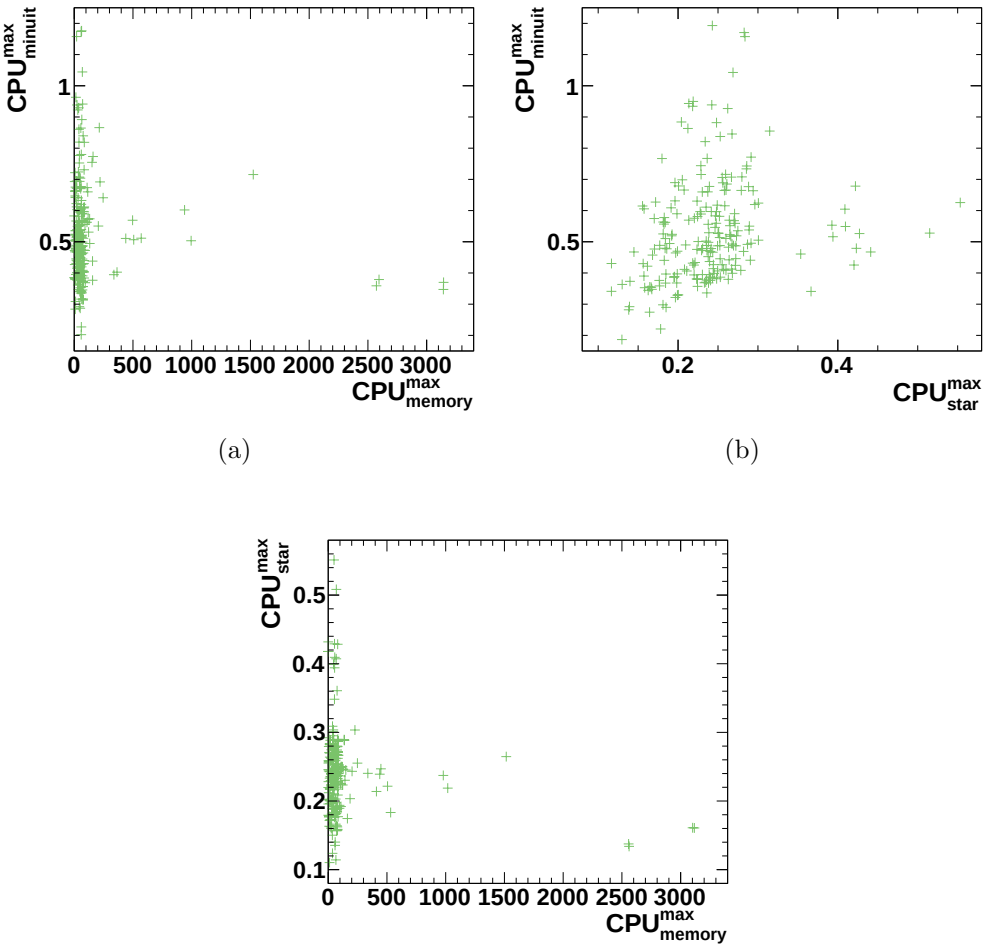
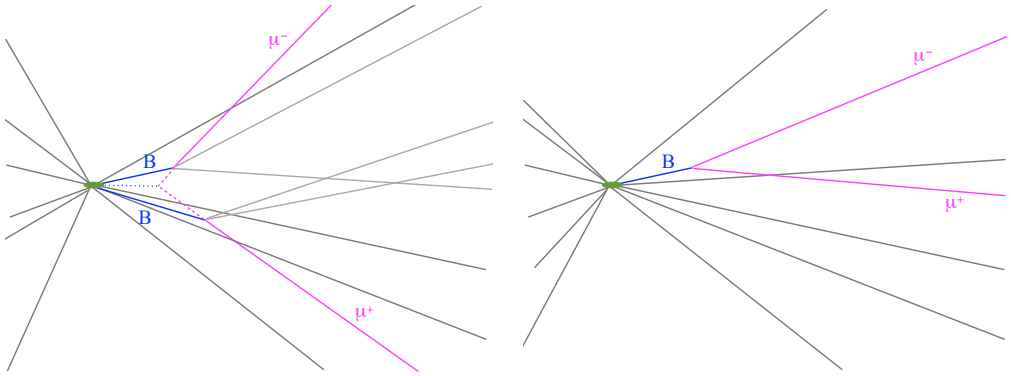
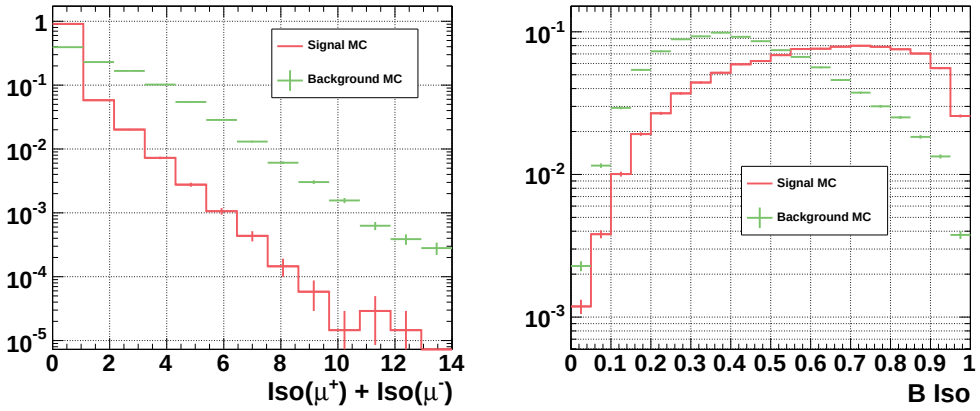


Figure 5.7: Comparison between the CPU used to find maxima by the Minuit and Memory (a), Star and Memory (b), and Star and Minuit (c) algorithms.



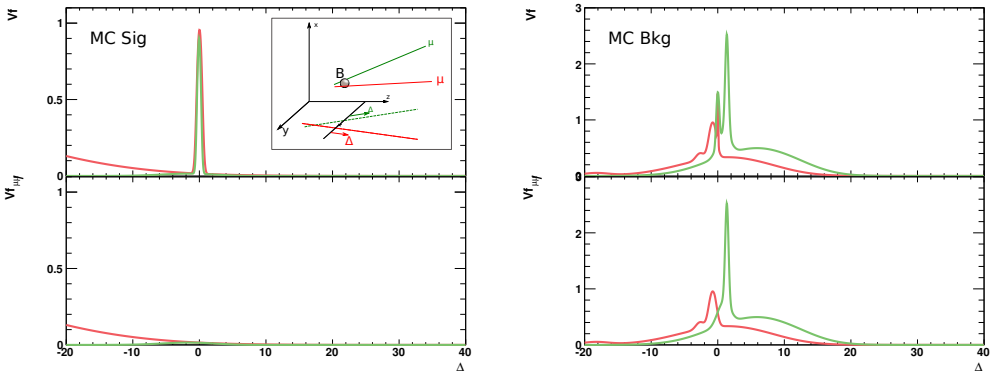
(a) (b)

Figure 5.8: Cartoon of typical combinatorial (a) and signal (b) candidates. The green ellipse is the primary vertex. The two pink lines are muons which originate from the decays of B mesons (blue). In the combinatorial case (a) their muons extrapolations (dashed pink) cross and make a B candidate vertex (dashed blue).



(a) (b)

Figure 5.9: Muon (a) and B (b) isolation variables. The red line represents the signal, the green the background, both from simulated data.



(a) (b)

Figure 5.10: Vertex function profiles along the muons tracks of a signal (a) and a combinatorial background (b) simulated candidate. The definition of the scan position coordinate is inlaid in frame (a).

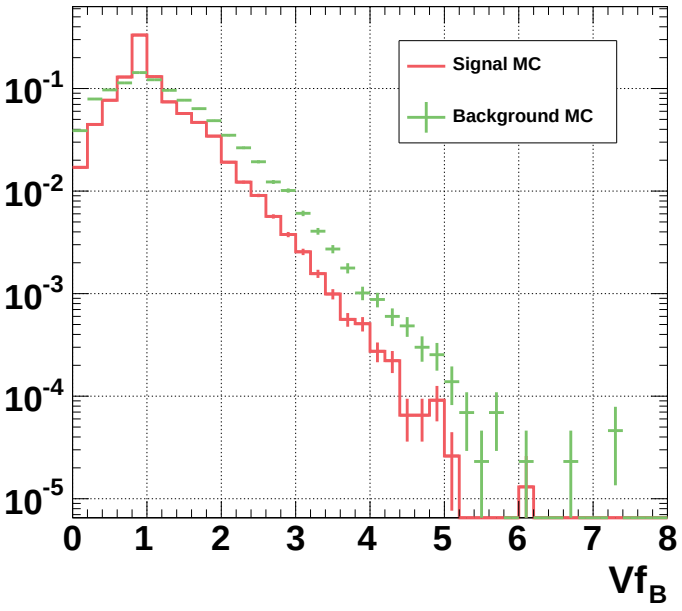
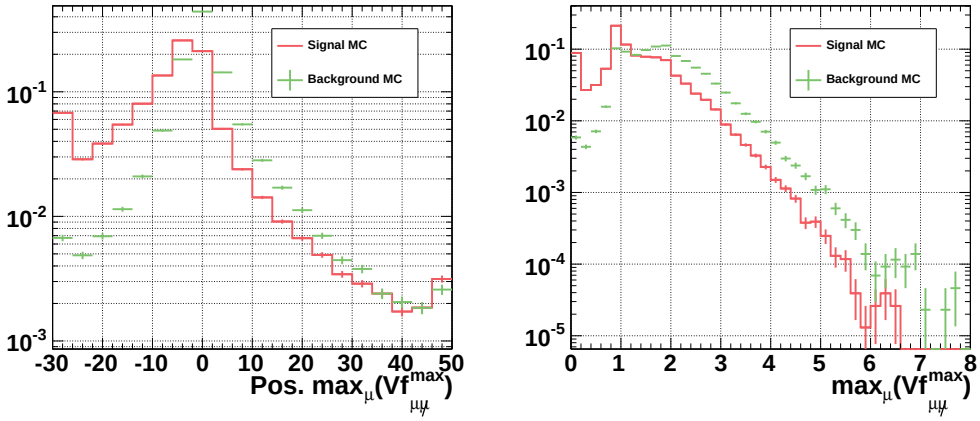


Figure 5.11: $V_{f,\mathrm{B}}$ distribution for the signal (red line) and combinatorial background (green crosses) from simulated data.



(a) (b)

Figure 5.12: Position (a) and $V_{f,\mu\mu}$ value (b) of the highest $V_{f,\mu\mu}$ maximum of the two muons tracks along them.

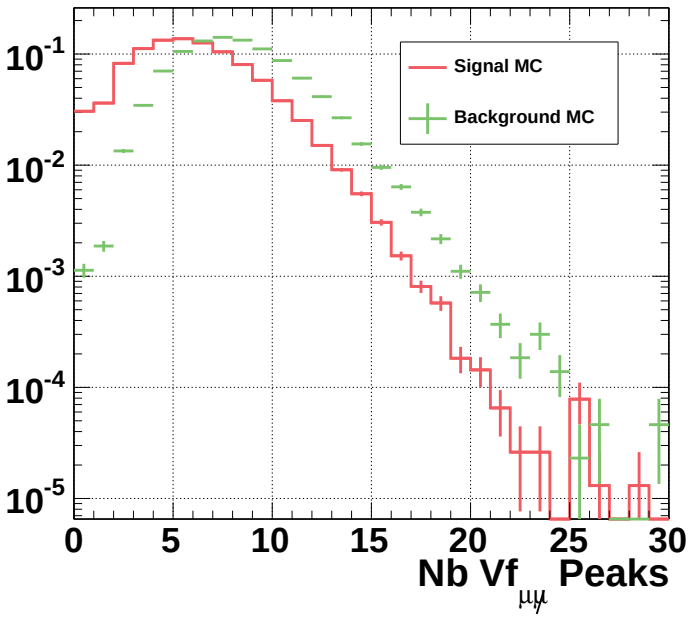


Figure 5.13: Number of $V_{f,\mu\mu}$ peaks along the two tracks.

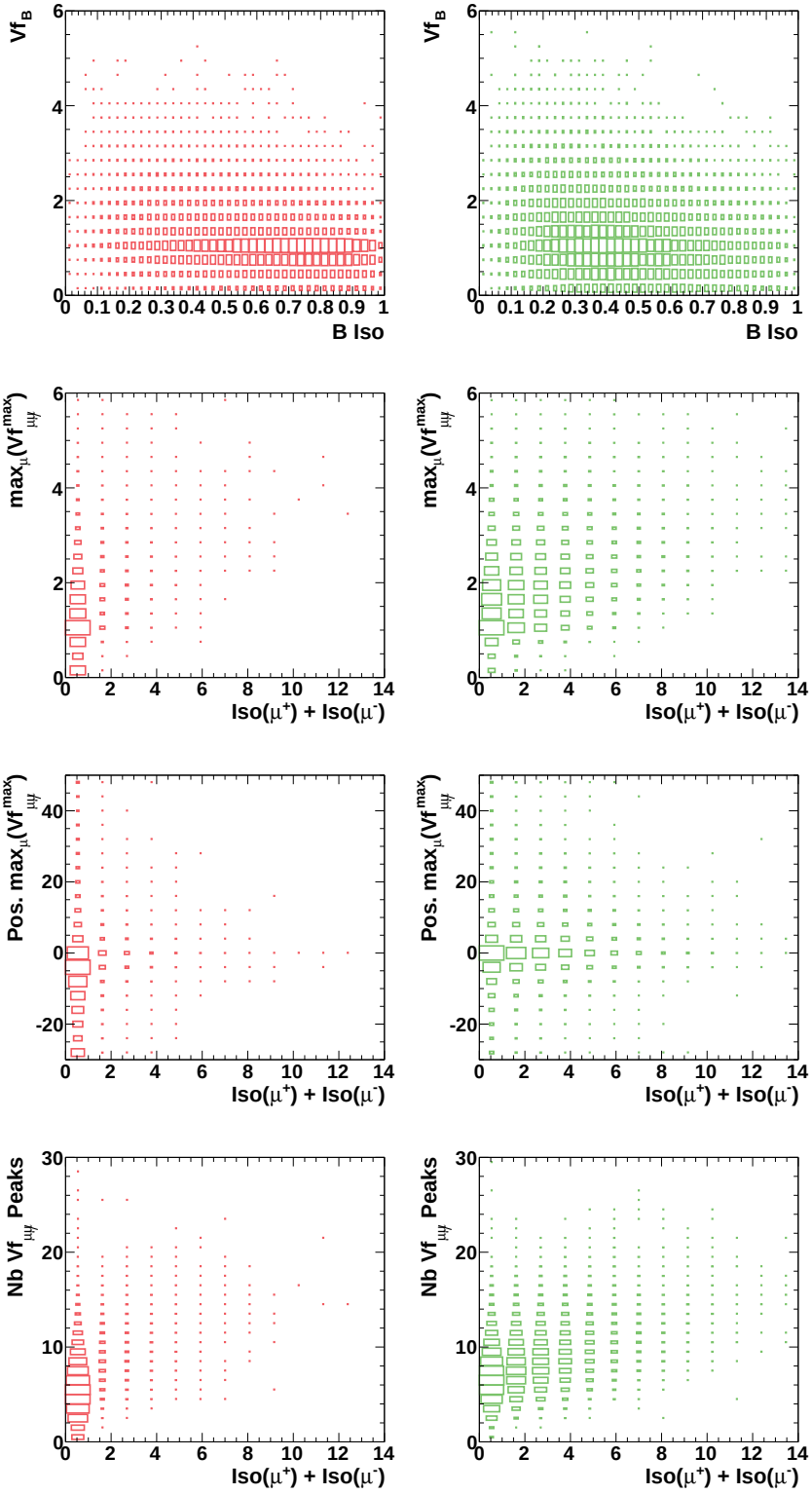


Figure 5.14: Scatter plots of the standard and V_f based B and muon isolations for the signal (red) and background (green).

Bibliography

[1] D. Jackson. *A topological vertex reconstruction algorithm for hadronic jets*. Nuclear Instruments and Methods in Physics Research, **A388** (1997) 1-2, 247–253.

[2] A. Strandlie et al. *Track and vertex reconstruction: From classical to adaptive methods*. Reviews of Modern Physics, **82** (2010), 1419–1458.

[3] D. Bailey et al. *The LCFIVertex package: vertexing, flavour tagging and vertex charge reconstruction with an ILC vertex detector*. Nuclear Instruments and Methods in Physics Research, **A610** (2009), 573–589.

[4] J. Seward, et al. *Valgrind 3.3-Advanced Debugging and Profiling for Gnu/Linux Applications* (2008).

[5] R. Fletcher et al. *Function minimization by conjugate gradients*. The Computer Journal, **7** (1964) 2, 149–154. URL `http://comjnl.oxfordjournals.org/content/7/2/149.abstract`.

[6] F. James et al. *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*. Computer Physics Communications, **10** (1975), 343–367.

Conclusions and Perspectives

The Standard Model (SM) of particle physics provides, overall, an accurate description of high energy physics. However, several theoretical and experimental issues led the particle physics community to conceive the SM as an effective low energy version of a more general theory. Hints for this new physics have been sought for decades, and searches continue today, as none were found.

This new high energy physics can be accessed through decays of heavy flavour mesons. Indeed, with an effective field theory, high energy processes responsible for a given heavy flavour meson decay can be, first, separated from the low energy processes and, second, classified in different types. This theoretical achievement allows to obtain accurate observable predictions for the SM and other new physics models. In this theoretical framework, experimental measurements of these observables are stringent tests of the SM and new physics models.

The $\mathsf{B}^0_{(s)} \to \ell^+\ell^-$ decays are among the most sensitive decays to new physics. They feature extremely small and well known SM contributions allowing new physics, if any, to completely drive their decay amplitudes. This extraordinary sensitivity to physics beyond the SM has motivated searches for these processes for three decades.

The Large Hadron Collider (LHC) provides the conditions to search for the $\mathsf{B}^0_{(s)} \to \ell^+\ell^-$ decays, and in particular the B^0_s modes. The LHCb detector was designed to operate at LHC and perform a physics program with $\mathsf{B}^0_s \to \mu^+\mu^-$ as one of the flagship measurements.

Soon after the first pp collision were recorded, the LHCb Collaboration was able to provide world best results on the $\mathsf{B}^0_{(s)} \to \mu^+\mu^-$ channels demonstrating the sensitivity of its detector. These results were improved as more luminosity was integrated and with $2.1\,\mathrm{fb}^{-1}$ the first evidence of $\mathsf{B}^0_s \to \mu^+\mu^-$ was reported with a significance of $3.5\ \sigma$ and a branching fraction of

$$\mathcal{B}(\mathsf{B}^0_s \to \mu^+\mu^-) \;=\; (3.2^{+1.4}_{-1.2}\ \text{stat.}\ ^{+0.5}_{-0.3}\ \text{syst.}) \times 10^{-9}. \qquad (5.4)$$

The two LHC general purpose experiments are also searching for these

modes, in particular CMS whose analyses reach a sensitivity comparable to LHCb. Therefore combining results allowed to significantly improve the sensitivity. At the time when this document was written, CMS had analysed only the data recorded in 2011 while the LHCb analysis was based on half of the data recorded in 2012 and therefore more sensitive than the CMS one. However the CMS results are still relevant. They allow, by combining them with the LHCb ones, to obtain an evidence with a slightly improved significance, 3.6 σ, and to constrain the $\mathsf{B}^0 \to \mu^+\mu^-$ branching fraction to be lower than 8.1×10^{-10} at 95% C.L..

These results point close towards the SM prediction and indicate that the understanding of new physics with the $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ channel will require a precision measurement. In wake of these results, a theoretical effort was made to clarify the definition of the observed quantities. These studies identified new extremely clean observables complementary to the decay branching fraction. One observable, $\mathcal{A}_{\Delta\Gamma}$, is of particular interest. Indeed, as long as it is not measured, a degree of freedom is left in the analysis. As a consequence, the experimental results depend on the value assumed for $\mathcal{A}_{\Delta\Gamma}$. This dependency can be addressed by providing branching fraction measurements as a function of $\mathcal{A}_{\Delta\Gamma}$. Since, theoretically, $\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-)$ and $\mathcal{A}_{\Delta\Gamma}$ have different correlations in different models, accounting for the $\mathcal{A}_{\Delta\Gamma}$ dependency of the experimentally measured branching fraction allows to reach higher sensitivity to some models. For instance, in the case of scalar models, some constraints on the Wilson coefficients can be improved by 30%.

In the future, the statistics available with the upgraded LHCb detector will allow to measure $\mathcal{A}_{\Delta\Gamma}$. On a shorter time scale, another theoretically clean observable, the ratio of the $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ and $\mathsf{B}^0 \to \mu^+\mu^-$ branching fractions, is accessible. Its precise measurement will probe the minimal flavour violation hypothesis, which is essential to understand the flavour pattern of the physics beyond the SM.

This pattern can also be understood by exploring other decay modes. In particular the study of the $\mathsf{B}^0_{(\mathsf{s})} \to \tau^+\tau^-$ decay would allow to understand if the couplings in physics beyond SM depend on the lepton family. Moreover, the heavy mass of the τ lepton removes suppression factors affecting some terms in the $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ decay amplitude and opens therefore new possibilities for physics beyond the SM to enter the $\mathsf{B}^0_{(\mathsf{s})} \to \tau^+\tau^-$ branching fraction. Searching for these modes is however challenging due to the short lifetime of the τ which must be reconstructed from their decay products involving necessarily neutrinos. In order to reconstruct these modes, and set a first upper limits on these processes, a topological inclusive event vertex reconstruction was developed. This method is based on a global function returning at each point the probability to have a vertex. The reconstruction of the decay ver-

tices is then equivalent in finding the maxima of this vertex function. A first attempt to use this method for $\mathsf{B}_\mathsf{s}^0 \to \tau^+\tau^-$ indicates that the signal efficiency could be improved by 21% with respect to a classical reconstruction, based on combinatorial trials, and reach 0.0175%. Finally, the vertex function appears to be, as well, a powerful tool to separate $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ signal from its combinatorial background and further applications, in other searches, can easily be foreseen.

Appendices

Appendix A

Variables Definitions

A.1 Track Isolation

For each of the muon candidates, a search is performed for long tracks (traversing all tracking detectors, with an impact parameter with respect to any primary vertex greater than 3 in units of significance, and of course excluding the other muon candidate), that can make a "good" vertex with the muon candidate. This requires the muon candidate and the given track to: make an angle < 0.27 rad, have a DOCA < 130 μm, and make a vertex whose distance PVdis (SVdis) to the primary (secondary) vertex satisfies: 4 cm $>$ PVdis > 0.5 cm (30 cm $>$ SVdis > -0.15 cm). The variables PVdis and SVdis are signed according to $z_{vtx}(\mu, tr) - z_{PV\ (\text{or}\ SV)}$. Furthermore, if we define $\alpha^{\mu+tr,PV}$ as the angle between the sum of the momenta of the muon and the extra track and the direction defined by the PV and the vertex reconstructed using the muon and the extra track candidates, then the sum of the momenta is required to satisfy

$$\frac{\left|\vec{P_\mu} + \vec{P_{tr}}\right| \cdot \sin(\alpha^{\mu+tr,PV})}{\left|\vec{P_\mu} + \vec{P_{tr}}\right| \cdot \sin(\alpha^{\mu+tr,PV}) + P_{T\mu} + P_{Ttr}} < 0.6 \tag{A.1}$$

where $P_{T\mu}$ and P_{Ttr} are the transverse momentum (with respect to the beam line) of the muon candidate and the extra track. The number of tracks that satisfy these conditions is used as a discriminating variable for each of the muon candidates.

A.2 B Isolation or "CDF Isolation"

The B isolation or "CDF isolation" (CDF definition [17]) I_{CDF} is defined as follows:

$$I_{CDF} = \frac{p_T(\mathsf{B})}{p_T(\mathsf{B}) + \sum_{tracks} p_T(tracks)} \tag{A.2}$$

where $p_T(\mathsf{B})$ is the B transverse momentum and the tracks used in the summation are those, excluding the candidates muons, for which $\sqrt{\delta\eta^2 + \delta\phi^2} < 1.0$, with $\delta\eta$ and $\delta\phi$ denoting respectively the difference in the pseudorapidity and of the ϕ coordinate between the track and the B candidate.

Appendix B

Latest Analyses Update

During spring 2013 the CMS and LHCb analyses were updated with the full 2012 dataset [1, 2]. Both analysis results were presented at the EPS conference together with their combination [3]. Section B.1 describes the main changes in the LHCb analysis with respect to the analysis presented in this thesis. Then, Section B.2 gives a brief overview of the CMS analysis. Finally the combined results are presented in Section B.3.

B.1 LHCb Analysis

The full LHCb dataset (1.0 fb^{-1} at $\sqrt{s} = 7\,\text{TeV}$ and 2.0 fb^{-1} at $\sqrt{s} =$ TeV) is analysed, the data used for the previous papers being fully re-analysed. The full $\sqrt{s} = 8\,\text{TeV}$ data are analysed together and their results combined with those obtained with the $\sqrt{s} = 7\,\text{TeV}$ data. The $\sqrt{s} = 8\,\text{TeV}$ data are reconstructed with improved algorithms and detector alignment parameters. Consequently, the invariant mass resolution is improved by around 5 MeV and the signal efficiency is increased. The multivariate classifier is re-optimised with new input variables and a larger combinatorial background simulated data sample. The corrections, introduced in Section 4.1, due to the non-trivial analysis lifetime acceptance and to the difference between the simulated data and SM signal lifetime distributions are implemented. Finally, f_s/f_d is updated [4] to account for the most recent measurements of B meson lifetime and of $\mathcal{B}(\mathsf{D_s^-} \to \mathsf{K^+K^-}\pi^-)$.

These modifications allows to improve the analysis sensitivity. With 50% more data, the previous analysis [5] would have reached an expected signal significance of 4.4σ. The improvements described previously allow to improve this significance up to 5.0σ. The CL_b corresponding to such a sensitivity is 5.7×10^{-9}. Hence, evaluating such a small probability requires

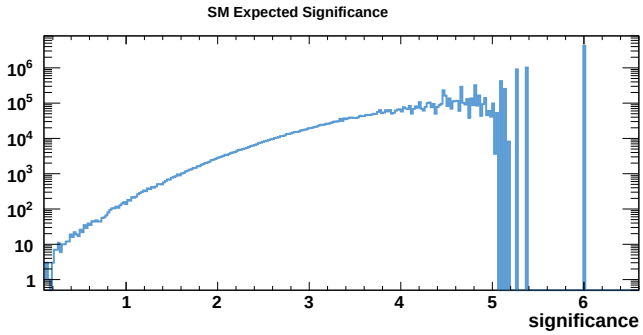


Figure B.1: Signal significance of 14 million of pseudo-experiments generated with a $B_s^0 \to \mu^+\mu^-$ rate compatible with the SM prediction.

several million of pseudo-experiments those generation has a non-negligible CPU cost. The uncertainties due to the finite number of pseudo-experiments should therefore considered carefully. For a given data pattern, the CL_b is evaluated as the integral of the background-only $-2\ln Q$ distribution above the data pattern $-2\ln Q$ value (see Section 3.1.9). When a dataset contains a significant number of signal events, its $-2\ln Q$ falls far away in the right tail of the background-only $-2\ln Q$ distribution. This distribution being obtained by generating background-only pseudo-experiments, the description of its tails is limited by the number of pseudo-experiments produced. Hence the integral to be computed to obtained the CL_b of the dataset considered is subject to a statistical error. In the worst case, no background-only pseudo pseudo-experiment has a $-2\ln Q$ value larger than the one of the dataset considered. In this case, a lower bound on the signal significance can only be given. Figure B.1 shows the signal significance distribution of 14 million of pseudo-experiments generated with a $B_s^0 \to \mu^+\mu^-$ rate compatible with the SM prediction. The expected sensitivity is defined as the median of this distribution which is $5.00^{+0.09}_{-0.06}$.

The data observed in the $B^0 \to \mu^+\mu^-$ mass window is compatible with the background-only ($CL_b = 5.6\%$). Hence an upper limit on the branching ratio is set. Table B.1 reports this upper limits as well as the expected ones and Figure B.2 gives the CL_s as a function of branching fraction hypothesis.

In the B_s^0 mass window the evidence seen in the previous analysis [5] is confirmed. The $B^0_{(s)} \to \mu^+\mu^-$ branching fractions are extracted by fitting simultaneously the invariant distributions in all the BDT bins of the 2011 and 2012 dataset. The significance of the $B_s^0 \to \mu^+\mu^-$ signal is 4.0σ with a

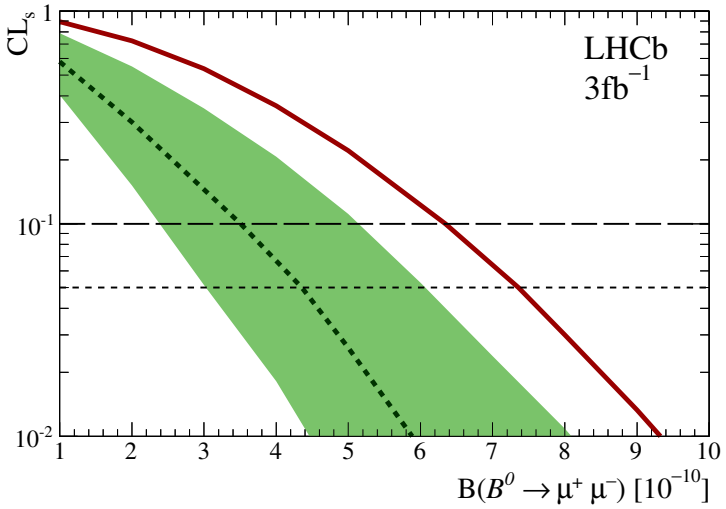


Figure B.2: CL_s as a function of the $B^0 \to \mu^+\mu^-$ branching fraction hypothesis. The black dotted line inlaid in the green band is the median of the CL_s distributions obtained for each branching fraction hypothesis under the assumption of the background-only and the band covers 34% on these distributions on each side of their median. The solid red curve is the observed CL_s. Branching fraction hypotheses corresponding to CL_s smaller than the threshold represented by the horizontal dashed (dotted) line are excluded at 90 (95)% C.L.

Table B.1: Expected limits for the background only (bkg) and background plus SM signal (bkg+SM) hypotheses, and observed limits on the $\mathsf{B}^0 \to \mu^+\mu^-$branching fraction.

	90 % C.L.	95 % C.L.
Exp. bkg	3.5×10^{-10}	4.4×10^{-10}
Exp. bkg+SM	4.5×10^{-10}	5.4×10^{-10}
Observed	6.3×10^{-10}	7.4×10^{-10}

rate of

$$\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-) \;=\; (2.9^{+1.1}_{-1.0}) \times 10^{-9}. \tag{B.1}$$

The $\mathsf{B}^0 \to \mu^+\mu^-$ branching fraction is measured at $(3.7^{+2.4}_{-2.1}) \times 10^{-10}$ but the signal has only a significance of 2.0σ (which is in agreement with the $\mathrm{CL_s}$ method results). The projection of mass fit in the most sensitive region is shown in Figure B.3.

B.2 CMS Analysis

The CMS collaboration analysed its full dataset (5 fb^{-1} at $\sqrt{s} = 7\,\mathrm{TeV}$ and $20\,\mathrm{fb}^{-1}$ at $\sqrt{s} = 8\,\mathrm{TeV}$). The $\sqrt{s} = 7\,\mathrm{TeV}$ and $\sqrt{s} = 8\,\mathrm{TeV}$ data are analysed separately and their results subsequently combined. The pile-up condition varies a lot over the data taking, therefore great attention was paid to remove pile-up dependencies in the analysis.

To avoid potential biases, the analysis is performed with the signal mass region kept blind until all selection criteria are chosen. A first BDT is built to reduce the background arising from misidentified hadrons. Then events are split in two categories called barrel and end-cap. The barrel category contains all candidates whose two muons have absolute pseudo-rapidity smaller than 1.4. The end-cap category gather events with at least one muon with a pseudorapidity larger than 1.4.

To extract signal events from the large remaining background ones, BDTs are built. As these BDT are built with data side-bands a special procedure was develop to avoid creating correlation with the candidate invariant mass. The data side-bands are split randomly into three sub-samples, a, b and c. The BDT to be used on sample c is trained on sample a and tested on sample b. As a results three BDT are built for each category (barrel, endcap, $\sqrt{s} = 7\,\mathrm{TeV}$, and $\sqrt{s} = 8\,\mathrm{TeV}$).

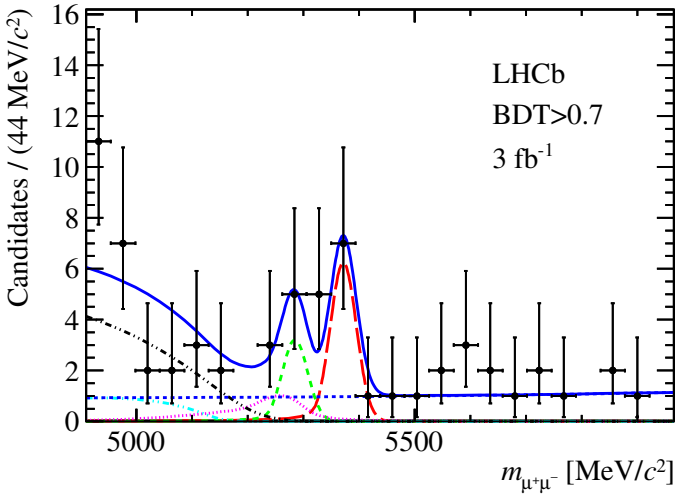


Figure B.3: Invariant mass distribution of the selected $B^0_{(s)} \to \mu^+\mu^-$ candidates (black dots) with BDT$>$ 0.7. The result of the fit is overlaid (blue solid line) and the different components detailed: $B^0_s \to \mu^+\mu^-$(red long dashed), $B^0 \to \mu^+\mu^-$ (green medium dashed), combinatorial background (blue medium dashed), $B^0_{(s)} \to h^+h'^-$(magenta dotted), $B^{0,+} \to \pi^{0,+}\mu^+\mu^-$ (light blue dot-dashed), $B^0 \to \pi^-\mu^+\bar{\nu}_\mu$ and $B^0_s \to K^-\mu^+\bar{\nu}_\mu$ (black dot-dashed).

These BDT are then used into two different ways called *1D-BDT* and *categorised-BDT*. In the 1D-BDT method, a selection criteria based on the BDT output is defined to maximise $S/\sqrt{S+B}$, where S and B are the numbers of signal and background event. In the categorised-BDT, the BDT output is used to define 12 categories with same expected signal yields. In both methods, all the categories (4 for the 1D-BDT method, and 4×12 for the categorised-BDT method) are combined to extract the $\mathsf{B}^0_{(s)} \to \mu^+\mu^-$ yields by a simultaneous fit to the invariant mass distribution and with the $\mathrm{CL_s}$ method to extract upper limits. The yields fitted are then converted into branching fractions by normalising to $\mathsf{B}^+ \to \mathsf{J}/\psi\mathsf{K}^+$. This normalisation employs the value of f_s/f_d measured by LHCb [6]. This quantities being dependent of the kinematics, a 5% systematics uncertainty is added to account for the different phase space between CMS and LHCb.

The categorised-BDT method is more sensitive to the signal and gives an expected significance for the signal of 4.8σ while the 1D-BDT method returns 4.7σ. These numbers show that the sensitivity of the CMS analysis is comparable with the LHCb one. The observed data exhibit a signal excess of a significance of 4.3σ using the categorised-BDT method[1] and the $\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-$ branching ratio is measured to be:

$$\mathcal{B}(\mathsf{B}^0_\mathsf{s} \to \mu^+\mu^-) \;=\; (3.0^{+1.0}_{-0.9}) \times 10^{-9}. \tag{B.2}$$

The $\mathsf{B}^0 \to \mu^+\mu^-$ branching fraction is measured at $(3.5^{+2.1}_{-1.8}) \times 10^{-10}$ with a significance of 2σ. The profile likelihood of the fit used to extract the branching fraction and the signal significance is shown in Figure B.4.

Even if the $\mathsf{B}^0 \to \mu^+\mu^-$ branching fraction is measured, its significance is only 2.0σ hence an upper limit on the branching fraction is also set, with the $\mathrm{CL_s}$ method, at

$$\mathcal{B}(\mathsf{B}^0 \to \mu^+\mu^-) \;<\; 11\ (9.2) \times 10^{-10} \text{ at } 95\ (90)\% \text{ C.L.}. \tag{B.3}$$

The $\mathrm{CL_s}$ as a function of branching fraction hypothesis is shown in Figure B.5.

B.3 CMS and LHCb Analysis Combination

While upper limits are more technically challenging to combine (see Section 3.5) branching ratios can easily be combined, at first approach, using averaging techniques. However, to extract combined signals significance more complicated method to combine likelihood functions are needed. A preliminary simple combination was therefore performed by the CMS and LHCb collaborations [3] after the individual results were presented.

[1] Surprisingly the 1D-BDT method returns a significance for the observed data of 4.8σ.

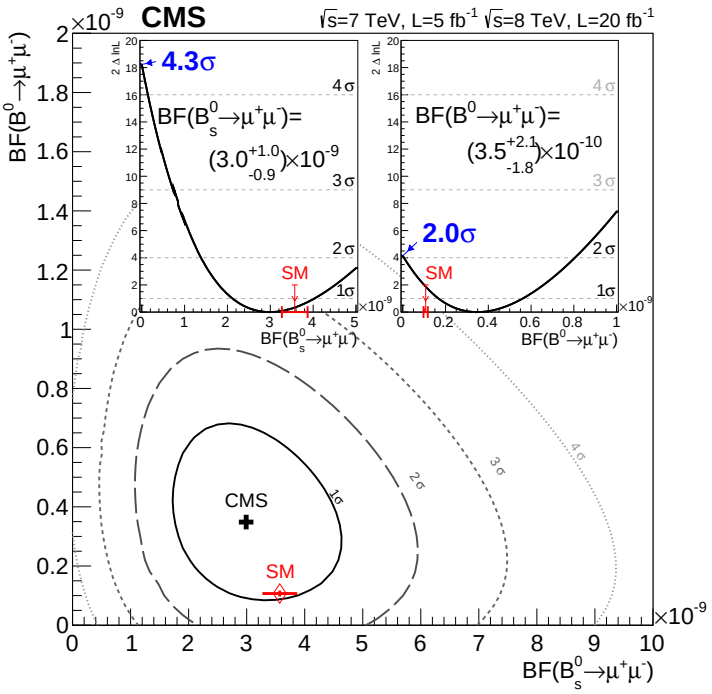


Figure B.4: Scan of the ratio of the joint likelihoods for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$. As insets, the likelihood ratio scan for each of the branching fractions when the other is profiled together with other nuisance parameters; the significance at which the background-only hypothesis is rejected is also shown.

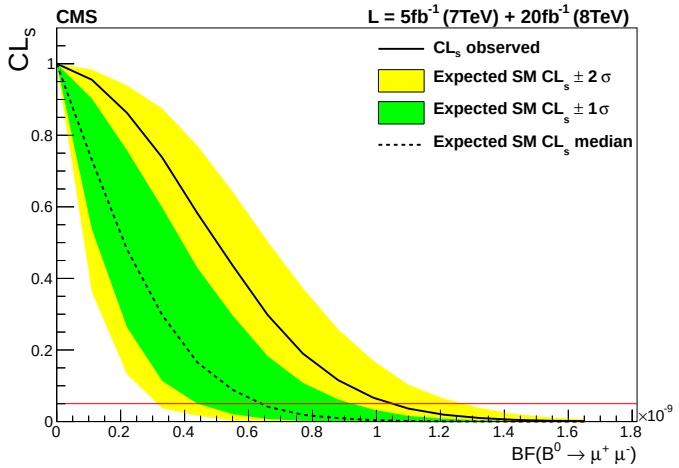
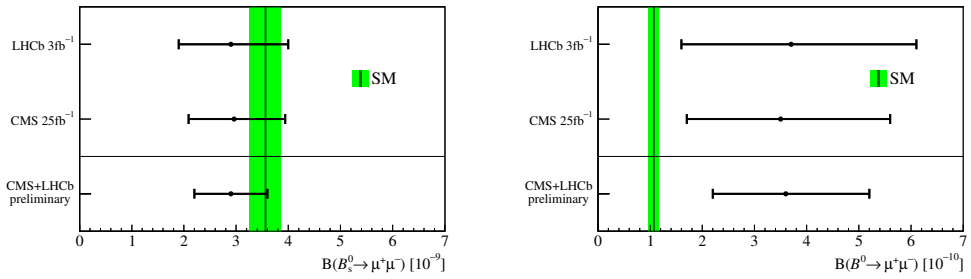


Figure B.5: CL_s as a function of the $B^0 \to \mu^+\mu^-$ branching fraction hypothesis. The black dotted line inlaid in the green and yellow bands is the median of the CL_s distributions obtained for each branching fraction hypothesis under the assumption of the background-only and the green (yellow) band covers 34% (68%) on these distributions on each side of their median. The solid black curve is the observed CL_s. Branching fraction hypotheses corresponding to CL_s smaller than the threshold represented by the horizontal red solid line are excluded at 95% C.L.

Since the CMS and LHCb analysis use common inputs, correlations must be handle carefully. The main relevant correlated input is f_s/f_d. Other inputs, like the branching fractions of $\mathrm{B}^+ \to \mathrm{J}/\psi\mathrm{K}^+$ and of the semi-leptonic b-decays, have negligible impact compared to this latter and their correlation are therefore ignored. In addition, the CMS result is updated with the more up-to-date value of f_s/f_d used by LHCb [4]. The averaged time-integrated branching fractions obtained are:

$$\begin{aligned} \mathcal{B}(\mathrm{B}^0_\mathrm{s} \to \mu^+\mu^-) &= (2.9 \pm 0.7) \times 10^{-9} \\ \mathcal{B}(\mathrm{B}^0 \to \mu^+\mu^-) &= \left(3.6^{+1.6}_{-1.4}\right) \times 10^{-10}. \end{aligned} \tag{B.4}$$

Although a proper likelihood combination is needed to extract the signals significance, it is clear that the $\mathrm{B}^0_\mathrm{s} \to \mu^+\mu^-$ signal significance exceeds 5σ while the $\mathrm{B}^0 \to \mu^+\mu^-$ one is below 3σ. Graphical comparisons between the individual results, the combined one, and the SM prediction is shown for $\mathrm{B}^0_\mathrm{s} \to \mu^+\mu^-$ and $\mathrm{B}^0 \to \mu^+\mu^-$ on Figure B.6.



(a) (b)

Figure B.6: Comparison of the latest CMS and LHCb results [1, 2], the combined value, and the SM prediction (vertical line) for (a) the time-integrated branching fraction $\mathcal{B}(\mathrm{B}^0_\mathrm{s} \to \mu^+\mu^-)$ and (b) $\mathcal{B}(\mathrm{B}^0 \to \mu^+\mu^-)$. The width of the vertical band represents the uncertainty in the SM prediction. The error bars represent the combined statistical and systematic uncertainties.

Bibliography

[1] CMS Collaboration. *Measurement of the* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *branching fraction and search for* $\mathsf{B}^0 \to \mu^+\mu^-$ *with the CMS Experiment.* Physical Review Letter, **111** (2013), 101804.

[2] LHCb Collaboration. *Measurement of the* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$ *branching fraction and search for* $\mathsf{B}^0 \to \mu^+\mu^-$ *decays at the LHCb experiment.* Physical Review Letter, **111** (2013) 10, 101805.

[3] CMS and LHCb Collaborations. *Combination of results on the rare decays* $\mathsf{B}_{(\mathsf{s})}^0 \to \mu^+\mu^-$ *from the CMS and LHCb experiments.* LHCb-CONF-2013-012 (2013).

[4] LHCb Collaboration. *Updated average* f_s/f_d *b-hadron production fraction ratio for 7 TeV pp collisions.* LHCb-CONF-2013-011 (2013).

[5] LHCb Collaboration. *First evidence for the decay* $\mathsf{B}_\mathsf{s}^0 \to \mu^+\mu^-$. Physical Review Letter, **110** (2013), 021801.

[6] LHCb Collaboration. *Measurement of the fragmentation fraction ratio* f_s/f_d *and its dependence on* B *meson kinematics.* Journal of High Energy Physics, **1304** (2013), 001.

Résumé

Les désintégrations $B^0_{(s)} \to \ell^+\ell^-$ sont supprimées d'hélicité et se produisent à travers des courants neutres changeant la saveur. Elles sont donc extrêmement rares dans le modèle standard (MS). La physique au-delà du MS pourrait donc modifier substantiellement la valeur des rapports d'embranchement prédits par le MS. Ces prédictions étant très précises, mesurer leur valeurs constitue un test incontournable du MS.

Les travaux présentés dans cette thèse ont contribué à l'établissement de la première évidence de la désintégration $B^0_s \to \mu^+\mu^-$ avec LHCb. Une méthode statistique a ete développée pour extraire et combiner les limites sur les rapports d'embranchement et calculer la signifiance des signaux. A partir de ces acquis statistiques, une méthode originale a été conçue pour re-optimiser les catégories dans lesquelles les candidats sont classés améliorant ainsi la sensibilité de l'analyse de 14%. D'autre part, une étude sur la phénoménologie des interférences entre le mélange de saveur des mésons B^0_s et leur désintégration a mis en évidence une subtile dépendance de l'analyse avec le modèle physique considéré. La prise en compte de cette dépendance permet de contraindre plus fortement la nouvelle physique. Enfin une méthode de reconstruction topologique inclusive des évènements a été développée et testée, notamment pour reconstruire les candidats $B^0_s \to \tau^+\tau^-$ avec $\tau^- \to \pi^-\pi^-\pi^+\nu_\tau$.

Abstract

The $B^0_{(s)} \to \ell^+\ell^-$ decays are helicity suppressed and process through flavour changing neutral currents. They are extremely rare in the standard model (SM). Thus physics beyond SM could modify their branching ratio up to an observable level. As the theoretical predictions are very precise, these modes are golden channels to probe the SM.

The works presented in this thesis have contributed to the establishment of the first evidence of the $B^0_s \to \mu^+\mu^-$ decay. A statistical method has been built to extract and combine limits on the decays branching fraction and to compute signals significance. On the basis of this work, an original method was developed to re-optimised the categories used to classified the candidates which improved the analysis sensitivity by 14%. A study of the interplay between the flavour mixing of the B^0_s and its decay demonstrated that the analysis results are model dependent. Accounting for this dependency allows to improve constraints on new physics. Finally, an inclusive topological reconstruction was developed and tested in particular to reconstruct $B^0_s \to \tau^+\tau^-$ candidates with $\tau^- \to \pi^-\pi^-\pi^+\nu_\tau$.